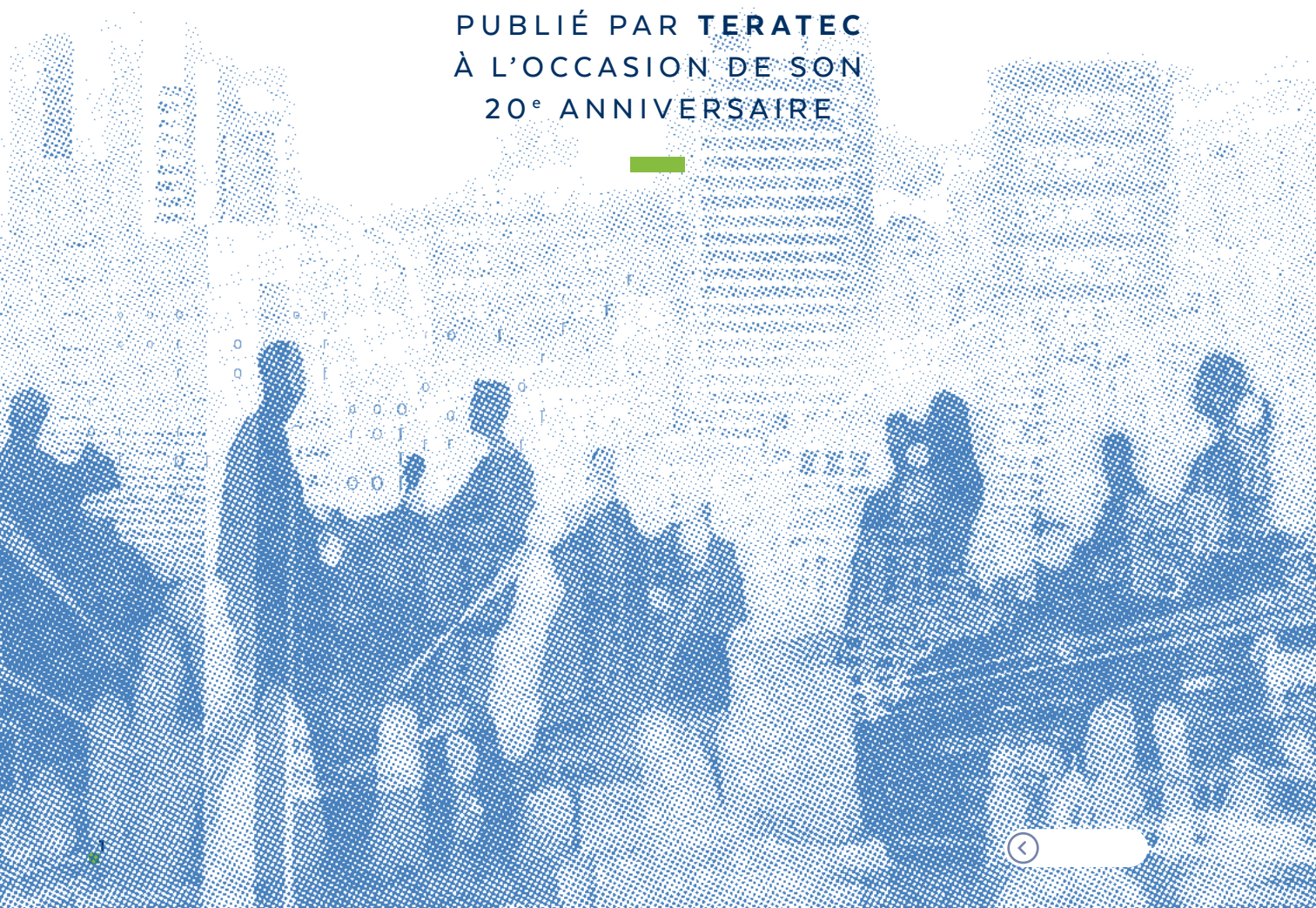
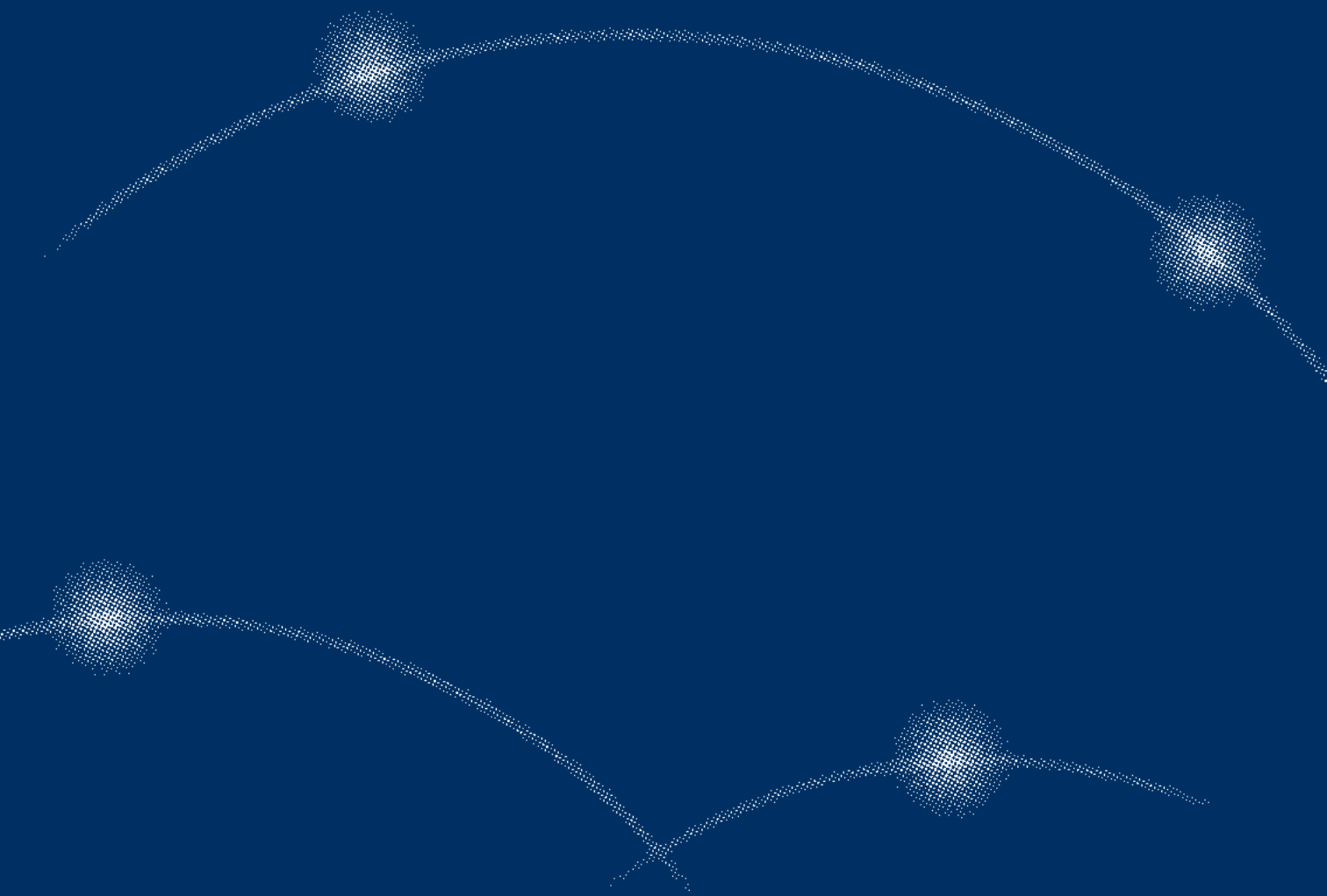




LES CLÉS DU NUMÉRIQUE

PUBLIÉ PAR TERATEC
À L'OCCASION DE SON
20^e ANNIVERSAIRE





LES CLÉS DU NUMÉRIQUE

PUBLIÉ PAR TERATEC
À L'OCCASION DE SON
20^e ANNIVERSAIRE



SOMMAIRE

INTRODUCTION

8 20 ANS APRÈS

Didier Besnard, Teratec

10 TERATEC A 20 ANS

Christian Saguez, Teratec

PERSPECTIVES

14 MAÎTRISER LES TECHNOLOGIES NUMÉRIQUES

Hervé Mouren, Teratec

16 L'IA POUR LES SCIENCES : DES PERSPECTIVES SCIENTIFIQUES FASCINANTES MAIS AUSSI QUELQUES POINTS DE VIGILANCE

Antoine Petit, CNRS

18 DU CONTRÔLE-COMMANDE À L'IA GÉNÉRATIVE : 40 ANS DE TRANSFORMATION NUMÉRIQUE DE L'INDUSTRIE

Patrice Caine, Thales

TECHNOLOGIES

22 INTRODUCTION : TECHNOLOGIES, DES MUTATIONS PROFONDES

Jean-François Prevéraud, Journaliste

23 ÉVOLUTION DES COMPOSANTS POUR LE CALCUL HAUTE PERFORMANCE : PERSPECTIVES À 5-10 ANS

Marc Duranton, CEA

27 L'INFORMATIQUE QUANTIQUE : VERS UNE NOUVELLE ÈRE DU CALCUL INTENSIF

Pascale Senellart, CNRS, Quandela

30 LES DÉFIS ÉNERGÉTIQUES DES ORDINATEURS QUANTIQUES

Olivier Ezratty, Quantum Energy Initiative

32 ÉVOLUTION DES ARCHITECTURES DE SUPERCALCULATEURS

Jean-Philippe Nominé, Patrick Carribault et Jean-Christophe Weill, CEA DIF

35 ARCHITECTURES HYBRIDES, MYTHES ET RÉALITÉS

Arnaud Bertrand, 3DS Outscale R&D

37 NVIDIA : OSER ÊTRE UN PIONNIER

Jensen Huang, NVidia

39 LES DÉFIS DU STOCKAGE : ÉVOLUER DANS UN PAYSAGE FRAGMENTÉ ET INNOVER À L'ÈRE DE L'IA

Jean-Thomas Acquaviva, DDN Storage

41 AVENIR DU STOCKAGE MASSIF DE DONNÉES POUR HPC ET IA

Gary Grider, Los Alamos National Laboratory

42 ADN, LA MÉMOIRE DU FUTUR

Marc Antonini, CNRS

43 VIRTUALISATION ET ORCHESTRATION

Guillaume Colin de Verdière, Gilles Wiber et Jean-Philippe Nominé, CEA

45 HPC ET EXASCALE : LA PUISSANCE À L'ÉPREUVE DE LA COMPLEXITÉ LOGICIELLE

Luc Giraud, Airbus CR & T, CERFACS et Brice Goglin, LaBRI

47 HPC ET INTELLIGENCE ARTIFICIELLE : INNOVATIONS ET CONVERGENCE

Gabriel Antoniu, Bruno Raffin et Olivier Beaumont, Inria

49 AGENCE DE PROGRAMME DANS LE NUMÉRIQUE, ALGORITHMES, LOGICIELS ET USAGES

Sophie Proust, Inria

50 L'IA CHANGE LA FAÇON DE CONCEVOIR

Herbert Taucher, Siemens

52 LE CALCUL AU CŒUR DES PRÉOCCUPATIONS DE L'AGENCE DE PROGRAMME ASIC

Jean-Philippe Bourgoin, CEA

USAGES

56 INTRODUCTION : USAGES, DE L'INDUSTRIE À L'HOMME

Jean-François Prevéraud, Journaliste

57 BIOLOGIE NUMÉRIQUE POUR EXPLORER ET PRÉDIRE LE VIVANT

Marjorie Domergue, Hervé Monod
et Carole Caranta, INRAE

60 SANTÉ & BIOLOGIE : MÉDECINE DE PRÉCISION ET INTELLIGENCE ARTIFICIELLE

Claude P. Bertrand, Servier

62 APPORTS DU HPC ET DE L'IA À LA SANTÉ NUMÉRIQUE : ENJEUX ET DÉFIS

Yves Vandenbrouck, CEA

64 NEUROSPIN : UN MARIAGE DE LA PHYSIQUE ET DU NUMÉRIQUE POUR SERVIR LES NEUROSCIENCES

Jean-François Mangin, CEA

66 LE JUMENT VIRTUEL HUMAIN : UNE AVANCÉE CRUCIALE POUR LA SCIENCE

Steven Levine, Dassault Systèmes

67 NUMÉRISER LE PATRIMOINE : DE LA PIERRE AU JUMENT NUMÉRIQUE

Florian Moreno, AGP

70 LE RÔLE ET L'APPORT DES TECHNOLOGIES HPC, IA ET QUANTIQUE FACE AUX DÉFIS CLIMATIQUES ET ENVIRONNEMENTAUX

Philippe Drobinski, CNRS,

72 LE NUMÉRIQUE AU CŒUR DES GÉOSCIENCES POUR UN MONDE EN MUTATIONS

Michaël Chelle, Romain Chassagne et
Simon Lopez, Bureau des Recherches
Géologiques et Minières (BRGM)

74 EXTENSION ET HYBRIDATION DU HPC DANS L'AÉRONAUTIQUE

Thierry Chevalier, Capgemini

76 TRANSFORMATION DIGITALE ET SIMULATION NUMÉRIQUE : L'ACCÉLÉRATEUR DES PROJETS VÉHICULES

William Becamel et Olivier Colmard,
Renault Group

78 HPC ET IA AU SERVICE DE LA SUPÉRIORITÉ NAVALE

Cédric Leblond, Naval Group

80 ENERGIE NUCLÉAIRE : L'INDISPENSABLE LIEN AU NUMÉRIQUE ET AU CALCUL INTENSIF

Patrick Blanc-Tranchant, CEA

**83 LES GRANDS ENJEUX DE
LA TRANSITION VERS LA NEUTRALITÉ
CARBONE ET LE RÔLE DE LA
TRANSFORMATION NUMÉRIQUE**

Alain Martin et Stéphane Tanguy, EDF
R&D

**85 IA HYBRIDE AU SERVICE DES VILLES
ET TERRITOIRES CONNECTÉS,
SOUTENABLES ET RÉSILIENTS**

Francisco Chinesta, Arts et Métiers
Sciences et Technologies, Dominique
Baillargeat, CNRS et Victor Champaney,
Duoverse

**88 ACCÉLÉRER LA DÉCOUVERTE
DE NOUVEAUX MATÉRIAUX : LA
CONVERGENCE DU HPC ET DE L'IA**

Jean-Yves P. Delannoy, Arkema

**90 IA, CALCUL INTENSIF ET QUANTIQUE
AU SERVICE DES FORCES DE
DEMAIN**

Emmanuel Chiva, Ministère des Armées
et Michaël Krajecki, Agence Ministérielle
pour l'IA de Défense

**93 LES TECHNOLOGIES DE CALCUL
HAUTE PERFORMANCE POUR LE
SECTEUR ASSURANCE-FINANCE**

Christophe Michel, Crédit Agricole CIB

**95 LE CLOUD CLÉ DE VOÛTE
DE L'INNOVATION, D'AUJOURD'HUI
ET DE DEMAIN !**

Yaniv Fdida, OVH Cloud

ASPECTS SOCIÉTAUX

**100 INTRODUCTION :
ASPECTS SOCIÉTAUX, DU BON
ET DU MAUVAIS**

Jean-François Prevéraud, Journaliste

**101 LA FORMATION FACE À LA VAGUE
DIGITALE : NOUVEAUX SAVOIRS
POUR NOUVEAUX DÉFIS**

Jean-Christophe Jouhaud, Cerfacs

**102 DES ENJEUX INÉDITS DANS
UN MONDE EN PROFONDE
MUTATION**

Thierry Damerval, Ministère de l'Europe
et des affaires étrangères

**104 ENCADRER L'INNOVATION
SANS LA FREINER : L'APPROCHE
EUROPÉENNE**

Alexandru Mateescu, Université Paris1
Panthéon-Sorbonne

**106 LE MONDE DE L'ENSEIGNEMENT
EST MOTEUR SUR LES NOUVELLES
TECHNOLOGIES NUMÉRIQUES**

Guillaume Gellé, Professeur des
Universités

**108 L'IA, ENTRE INNOVATION ET
RESPONSABILITÉ, LE SOMMET
DE PARIS A MARQUÉ UNE ÉTAPE
HISTORIQUE**

Anne Bouverot, Envoyée Spéciale
du Président de la République pour
le Sommet pour l'Action sur l'IA

20 ANS APRÈS

Didier Besnard

Président, Teratec

Teratec a 20 ans. Grâce à l'implication de ses membres, cette association a largement contribué à la souveraineté de la France en simulation haute performance, que ce soit en offre de technologie et en usages industriels. Teratec est au cœur d'un écosystème, associant fournisseurs de technologies, utilisateurs des technologies du HPC et institutions de recherche. Cela n'a été rendu possible que grâce à une articulation forte entre ses membres industriels et ses membres académiques: la science est dans le temps long, les applications et l'innovation dans le temps court. Teratec contribue à marier les deux.

Teratec est ainsi un lieu d'échanges et de génération d'initiatives et de projets, portés au niveau national et européen, unique en son genre grâce à la science et l'innovation générées par ses membres. L'écosystème de Teratec, celui de la simulation haute performance et des données massives, incluant le quantique et l'IA, est en évolution rapide. Celle-ci est portée pas un plus grand parallélisme entre le progrès scientifique et l'émergence d'applications industrielles et sociétales: la facilité de création de start-up pour les chercheurs, les initiatives nationales et européennes de soutien à l'innovation en sont des moteurs essentiels. De plus, la dynamique scientifique elle-même des différentes branches de la simulation haute performance permet aussi l'émergence d'outils plus performants. Ce sont autant de défis et d'opportunités, en science et technologie, d'une part, et en apport aux entreprises de l'écosystème, d'autre part.

DANS UN ÉCOSYSTÈME EN ÉVOLUTION RAPIDE, TERATEC VOIT SA MISSION RENFORCÉE

A l'aune de ces remarques, Teratec peut et doit continuer à s'insérer à la fois dans ces dynamiques scientifiques à l'œuvre:

- Beaucoup des applications industrielles envisagées par les membres de Teratec sont des applications critiques. La précision et la fiabilité des calculs doivent donc être assurées. Si l'hybridation entre simulation physique et IA offre des possibilités d'accélération à la fois de la modélisation et des simulations numériques elles-mêmes, il faut pouvoir en démontrer le respect des lois physiques, et avoir une mesure des incertitudes du calcul. Par ailleurs, l'utilisation maîtrisée de l'IA pour traiter des données massives obtenues dans tous les domaines, scientifiques et/ou industriels, vient compléter le triptyque fondamental de la Simulation, modélisation, simulation numérique, et validation expérimentale. Les applications sont sans limite: modélisation par les données, simulateurs et données synthétiques, robotique... C'est un axe fort pour Teratec.
- Les technologies quantiques évoluent de plus en plus rapidement. L'initiative Teratec autour du calcul quantique (TQCI), lancée il y a quelques années, s'intègre naturellement dans la forte dynamique actuelle. L'élargissement des participants, l'intégration des start-up du domaine, la mise en place par le SGPI des Maisons du Quantique, sont autant d'actions à amplifier.

- L'IA est un puissant moteur de productivité et d'aide aux utilisateurs non spécialistes d'une technologie. Elle permettra un accès plus facile à la simulation haute performance si des outils simples peuvent être développés sur la base d'IA spécialisées. C'est aussi le cas pour l'utilisation de technologies IA par des chercheurs non spécialistes du domaine pour leur propre application. Il y a donc un enjeu à la fois d'outils, et de formation associée.

OPÉRATIONNELLEMENT, TERATEC VA AMPLIFIER SA CONTRIBUTION À L'ÉCOSYSTÈME DE LA SIMULATION.

Quels en sont les défis ?

- Améliorer les passerelles entre science et innovation au sein de l'association, en renforçant l'articulation entre les membres industriels et les membres académiques.

Des actions dans ce sens viennent d'être entreprises avec les membres actuels, mais l'élargissement à de nouveaux membres académiques est à faire.

Travailler avec les start-up est aussi dans le droit fil de cet objectif. Bon nombre sont issues des recherches actuellement menées dans les organismes de recherche. Le soutien aux start-up doit se renforcer, notamment celles développant de nouvelles technologies matérielles de calcul intensif, du quantique, ou pouvant être spécialisées à l'IA.

- Il faut également élargir l'empreinte de Teratec en renforçant les liens, à l'échelle européenne, avec les acteurs de la science et de la technologie, et les acteurs institutionnels. L'échelle européenne est un gage de plus grand poids et d'efficacité. Au point de vue technique, des résultats prometteurs ont déjà été obtenus par l'entremise de projets européens.

Il faut en capitaliser les résultats déjà obtenus, en diffusant et rendant pérennes les outils et/ou actions de formation déjà développés, et en les utilisant au profit des membres de l'association.

- Teratec, de par ses membres, est au cœur de l'écosystème de la Simulation haute performance. Pour autant, il lui faut encore mieux s'intégrer dans les initiatives régionales et nationales de recherche et d'innovation. A ce titre, elle aura à lancer ou à contribuer à de nouvelles initiatives de souveraineté initiées pas ses membres.

- Enfin, les points développés plus haut militent fortement pour le développement d'un axe de formation et d'accompagnement, en s'associant à des acteurs scientifiques et techniques le faisant déjà. A titre d'exemple, il s'agit de promouvoir l'accès des technologies de la simulation haute performance par des outils simples et conviviaux basés sur l'IA au sein des filières industrielles, jusqu'aux PME/PMI ; Il s'agit aussi de faciliter l'accès aux outils d'aide au développement de codes basés sur l'IA pour les non-spécialistes.

TERATEC CONJUGUÉ AU FUTUR

Avec ses actions scientifiques et techniques, le renforcement des interactions entre membres en réponse à leurs besoins, l'élargissement de son empreinte régionale, nationale et européenne, et ses actions de formation, Teratec renforcera son caractère unique et répondra aux exigences d'un écosystème en évolution rapide.

Vingt ans après, Teratec reste à l'unisson des évolutions en cours dans son domaine. Les perspectives sont enthousiasmantes, les défis existent. L'effort est à continuer. ■

TERATEC A 20 ANS

Christian Saguez

Co-fondateur et président d'honneur de Teratec

Depuis sa création en 2005, le calcul haute performance (HPC) est devenu un élément stratégique majeur. Dans les années 2000, la France ne maîtrisait plus ni les technologies matérielles ni logicielles du HPC. Le programme Simulation du CEA révéla cette fragilité, mettant en danger notre souveraineté, notre maîtrise scientifique et notre développement économique. Le numérique allait devenir la clé pour concevoir, innover, produire durablement et exploiter les innovations. Peu de responsables, notamment politiques, en avaient conscience, et il fallut livrer une bataille intense pour convaincre, alors qu'un célèbre cabinet de conseil en niait l'importance.

À l'initiative du CEA, un petit groupe d'industriels créa Teratec, association loi 1901, lançant une dynamique qui a remis la France dans le concert des grandes nations numériques et instauré un modèle unique en Europe.

« Ce n'est pas parce que les choses sont difficiles, que nous n'osons pas, c'est parce que nous n'osons pas que les choses sont difficiles » a dit Sénèque, alors nous avons osé, malgré les nombreux obstacles et oppositions et aujourd'hui Teratec fête ses 20 ans.

CRÉER UNE DYNAMIQUE EN FÉDÉRANT OFFREURS DE TECHNOLOGIES, UTILISATEURS ET CENTRES DE RECHERCHE

Dès sa création, des principes essentiels ont été posés :

- Associer offreurs de technologie, utilisateurs industriels et centres de recherche dans de grands projets collaboratifs.

- Développer une philosophie de co-design en réunissant matériel, logiciel et architectures au sein d'équipes pluridisciplinaires.

- Développer les usages dans tous les secteurs économiques et les diffuser auprès des entreprises de toute taille, y compris les PME.

Les usages du HPC issus de la modélisation mathématique par EDP (Équation aux dérivées partielles) et grands systèmes algébro-différentiels concernaient surtout les grandes industries manufacturières, énergie, aéronautique, automobile, avec de grands codes de calcul basés sur les éléments finis et des architectures parallèles.

Aujourd'hui, les calculateurs, enrichis de GPU et autres accélérateurs, offrent des puissances accrues. Se sont développées des plateformes de calcul (cloud, edge) et de nouveaux modèles économiques comme le SaaS (Software as a Service). Les data centers se sont multipliés et de nouveaux défis sont apparus : consommation énergétique, précision numérique, sécurité, algorithmie multidisciplinaire. Le secteur a ainsi profondément évolué en une décennie.

Dans ce contexte, Teratec est aujourd'hui un acteur majeur en Europe et doit poursuivre son action par sa capacité à fédérer, sa force de proposition et son travail de prospective dans un secteur en pleine transformation.

DES ÉVOLUTIONS EXTRÊMEMENT RAPIDES AUTOUR DES TECHNOLOGIES DU HPC/IA/ QUANTIQUE

De nombreuses évolutions technologiques et sociétales marqueront un monde largement numérique et parallèle. Parmi elles, on peut souligner les points suivants :

- Le rôle croissant des données, grâce à leur disponibilité massive. Elles permettent d'associer aux modèles mathématiques traditionnels (issus de la physique) des outils comme les réseaux neuronaux et le Deep Learning, ouvrant la voie à une approche globale de problèmes complexes. Cela a généralisé le concept de jumeau numérique. Le traitement de ces données, combiné à la puissance de calcul, alimente aussi l'essor de l'IA générative, dont les usages devront être maîtrisés.
- L'extension continue des usages numériques à tous les secteurs industriels, professionnels et de la vie quotidienne. Cela concerne des enjeux cruciaux comme la mobilité, la ville durable, l'agriculture et l'alimentation, la santé et la biologie (nouveaux médicaments, thérapies, médecine personnalisée), mais aussi le climat, l'environnement, la biodiversité, sans oublier des domaines tels que le cinéma, les jeux vidéo, la culture, le commerce, la mode, le textile et les réseaux sociaux.
- L'arrivée de nouvelles technologies, pouvant être de rupture, qui apportera sans doute de nouvelles capacités d'analyse et de découverte, tel le stockage sur molécules et ADN, le calcul quantique, l'ordinateur neuromorphique... Ces technologies associées aux technologies actuelles vont générer à tous les niveaux, matériel, logiciel, algorithmie. Une nécessaire hybridation de toutes les approches matérielles, logicielles, algorithmiques. Ceci nécessite un effort de recherche continu sur le temps long.

UN ENJEU FONDAMENTAL POUR LA FORMATION ET LES MATHÉMATIQUES

Face à ces changements, notre société doit être en mesure de les maîtriser, de les comprendre et d'en faire un usage positif, tout en contrôlant les dérives inévitables. La formation initiale et continue constitue un enjeu majeur : disposer d'un nombre suffisant de

techniciens et d'ingénieurs pour garantir le développement, créer valeur et emplois, mais aussi permettre à chacun un usage raisonné afin d'éviter les dérives toujours liées à de telles innovations. Toute innovation apporte des opportunités et des risques, mais leur rejet serait à long terme très dommageable pour notre société.

Dans ce contexte, il faut insister sur le rôle majeur des Mathématiques. Face à la baisse continue du niveau des étudiants, il est urgent de mettre en place des mesures pour réhabiliter cette discipline fondamentale, en menant une large campagne de promotion, afin de la rendre plus attractive avec les nombreux métiers passionnants qui en découlent, mais aussi pour mieux comprendre ses usages. C'est indispensable pour répondre aux défis présents et futurs. Comme le disait Einstein, « un homme qui a cessé de s'émerveiller a cessé de vivre » : continuons donc à nous émerveiller des découvertes, à soutenir la recherche et à rester confiants en l'avenir.

UNE NÉCESSITÉ ET UNE CHANCE POUR LA FRANCE ET POUR L'EUROPE

Les technologies numériques sont et resteront un élément fondamental pour l'innovation, la compétitivité et la création de valeur dans tous les secteurs de l'économie. Elles interviendront dans toutes nos activités professionnelles et de la vie quotidienne. La France dispose de nombreux atouts, mais doit agir vite pour maintenir une position d'excellence face à une concurrence mondiale croissante, en sachant former les spécialistes grâce à un enseignement scientifique et mathématique renouvelé et adapté. Cela doit s'inscrire dans un processus continu, sans céder aux effets de mode et avec une forte dimension européenne.

Dans ce cadre, Teratec doit continuer de jouer un rôle de premier plan comme pôle d'excellence, d'analyse et de prospective, en soutenant les initiatives régionales, nationales et européennes indispensables pour assurer souveraineté et compétitivité, et en garantissant leur promotion auprès de tous les acteurs. Sachons saisir cette chance et restons confiants en un avenir plein d'innovations et de découvertes où le numérique apportera encore bien des nouveautés aujourd'hui insoupçonnées. ■





PERSPECTIVES



MAÎTRISER LES TECHNOLOGIES NUMÉRIQUES

Hervé Mouren

Directeur, Teratec

Le développement très rapide des technologies numériques, depuis le calcul intensif et la simulation jusqu'à l'intelligence artificielle, et demain le quantique, bouleverse très profondément nos façons de travailler et de concevoir produits et services. Dès 2005, les grands industriels de l'aéronautique et de l'énergie avaient mesuré la transformation profonde que cela ouvrait en matière de conception et de simulation.

Teratec a aujourd'hui quatre-vingts membres et s'est développé sous la forme d'un écosystème varié, très complémentaire, regroupant utilisateurs, fournisseurs, enseignants et chercheurs. Son objectif est d'assurer à ses membres la maîtrise des technologies numériques et de faciliter leur diffusion et leur déploiement. Ces deux aspects, maîtrise et diffusion, sont clés parce que le numérique permet de concevoir et de réaliser des produits ou des services impossibles à envisager précédemment et qu'il devient de fait un des facteurs essentiels de compétitivité industrielle. Son champ d'application s'élargit aujourd'hui à la plupart des domaines, tous les secteurs sont maintenant concernés et progressivement toutes les tailles d'entreprises.

Teratec accompagne et contribue activement depuis 20 ans à la généralisation du recours au calcul intensif et à la simulation, tandis que pendant cette période, la puissance de calcul des superordinateurs a été multipliée par 1 million : le passage du Téraflopp à l'Exaflupp nous permet de traiter en 2025 des sujets dont on

n'avait pas idée en 2005. De plus, l'intelligence artificielle et le quantique qui se développent à grande vitesse ouvrent une nouvelle phase exceptionnelle de développements la plupart complètement nouveaux. C'est cela que nous devons préparer.

Pour aider à la prise en compte par un public de plus en plus large, Teratec organise chaque année le Forum Teratec, qui est le grand moment d'échange en France entre chercheurs et industriels, fournisseurs et utilisateurs.

Teratec participe par ailleurs de façon très active à la nouvelle dynamique européenne qui s'est créée sur ces sujets. Nous sommes une des parties prenantes du programme européen EuroHPC qui bénéficie d'un budget de 7 milliards d'euros sur 6 ans avec notamment la commande de deux machines de niveau exascale, dont la première, installée en Allemagne, est réalisée par des acteurs français et dont la seconde est prévue l'année prochaine à Bruyères-le-Châtel, dans l'Essonne. Toujours dans le cadre de ce programme européen, Teratec participe à de nombreux projets et, en particulier, opère avec ses partenaires un centre de compétence français pour les technologies numériques, qui couvre tous les champs du HPC, de l'IA et du quantique et qui fournit support et formations à tous les utilisateurs du monde industriel et du monde de la recherche.

Aujourd'hui, le numérique touche tous les domaines industriels et s'est imposé comme une composante essentielle de notre développement économique. Il est devenu un des principaux différenciateurs de compétitivité

et d'efficacité des entreprises dans tous les domaines et à tous les niveaux, national, européen et mondial.

La France et l'Europe ne mesurent pas encore pleinement l'enjeu stratégique que représentent la montée en compétences scientifiques et la maîtrise des technologies critiques. D'autres régions du monde ont réalisé en dix ans des avancées majeures. Nous avons les moyens de faire de même, avec un bon niveau de départ : excellence scientifique, qualité de la formation. Ce potentiel reste trop méconnu ; par exemple, certains des outils d'apprentissage-machine les plus utilisés au monde ont été développés dans nos laboratoires.

COMPÉTENCES ET FORMATION FERONT LA DIFFÉRENCE

Le numérique évolue très vite, faisant de la formation un enjeu stratégique. Les organisations qui investissent dès aujourd'hui dans la maîtrise de ces technologies bénéficieront d'un avantage compétitif. Ceux qui les maîtriseront le mieux seront les leaders de demain. Dans la formation initiale, il faut non seulement développer des cursus dédiés, mais aussi préparer tous les futurs ingénieurs à leur usage. En parallèle, nos équipes actuelles de recherche et d'ingénierie doivent intégrer ces outils sans attendre l'arrivée de spécialistes. La formation continue est donc essentielle. Le numérique étant une course de vitesse, il faut accélérer la montée en compétences des ingénieurs en poste, notamment en R&D. C'est pour Teratec un enjeu majeur et une priorité.

Un autre enjeu concerne la question très difficile de la protection des données et plus largement de la cybersécurité. En effet, la protection des données individuelles, qu'elles soient administratives, financières ou de santé est un sujet majeur, au cœur des préoccupations des entreprises et de l'État et de toutes les organisations publiques ou privées. Nous évoluons dans un monde de plus en plus digitalisé où on recense une augmentation continue des attaques informatiques et de la cybercriminalité. Plus que jamais, il est essentiel de pouvoir garantir la confidentialité et l'intégrité des informations partagées en ligne. C'est le risque majeur de la période qui vient auquel nous devons être très attentifs en permanence.

Il y a par ailleurs un enjeu énergétique très important qui apparaît avec les investissements colossaux prévus

pour l'intelligence artificielle générative. Cette fuite en avant risque d'avoir des conséquences lourdes qu'on a du mal à mesurer aujourd'hui.

Au cours des dernières décennies, la digitalisation a permis l'essor du courrier électronique, du télétravail, du commerce électronique, mais également une plus large diffusion de la culture au travers de plateformes digitales, la musique et la vidéo en étant deux bons exemples. Aujourd'hui, toujours plus de secteurs et d'industries capitalisent sur ces savoir-faire et ces compétences : l'agroalimentaire, la santé, la biologie, le transport, la finance, l'assurance, la culture, l'environnement et le climat, et tout le domaine des sciences humaines. La digitalisation va permettre des avancées très spectaculaires dans le domaine de la santé. Les technologies numériques ouvrent des possibilités nouvelles notamment pour soigner des pathologies que nous avons du mal à traiter jusque-là ou pour gérer des risques d'épidémie. Les grands progrès dans le domaine de la santé sont devant nous, c'est un enjeu magnifique.

Teratec et ses membres sont en veille constante sur les évolutions en cours, avec le développement ultrarapide de l'intelligence artificielle, mais aussi du quantique avec des premières applications industrielles qui ne vont plus tarder à voir le jour et qui seront pour beaucoup révolutionnaires, car conçues sur des approches très différentes. C'est en 2018, il y a sept ans, que nous avons lancé la Teratec Quantum Computing Initiative (TQCI) pour nous préparer à l'arrivée du calcul quantique. Cette initiative nous a permis de mener nombre de projets auxquels sont associés maintenant de nombreux pays. C'est une très bonne illustration de ce principe de co-création (co-design) associant fournisseurs et utilisateurs, qui, dans un domaine comme le nôtre s'avère particulièrement efficace.

Sur tous ces sujets, la France a réussi à maintenir une position de premier plan en matière de recherche et de formation. Dans le contexte géopolitique international actuel, il est essentiel de préserver cette place, voire de la renforcer en misant sur la formation et de donner la possibilité à nos ingénieurs et nos chercheurs d'intégrer le plus en amont possible les développements technologiques en cours. C'est un des objectifs de Teratec, aujourd'hui et pour les années à venir. ■

L'IA POUR LES SCIENCES : DES PERSPECTIVES SCIENTIFIQUES FASCINANTES MAIS AUSSI QUELQUES POINTS DE VIGILANCE

Antoine Petit
PDG, CNRS

L'intelligence artificielle (IA) transforme le monde et notamment les sciences avec la promesse excitante d'accélérer la découverte scientifique mais aussi avec des risques dont il faut avoir pleinement conscience.

L'IA n'est pas née avec ChatGPT. Voilà près de 70 ans, en 1956, que s'est tenu le « *Dartmouth Summer Research Project on Artificial Intelligence* », considéré comme l'événement fondateur du domaine de l'intelligence artificielle.

Depuis longtemps donc des communautés scientifiques travaillent en mathématiques appliquées, en informatique et sciences des données, sur le raisonnement et l'apprentissage machine, sur le traitement automatique des signaux sur l'automatique, la robotique ou encore l'interaction humain machine. Et depuis presque aussi longtemps, des travaux visent à utiliser l'intelligence artificielle (sans que le terme soit nécessairement utilisé) en lien avec d'autres disciplines. Le traitement des données en physique des particules (issues notamment du CERN) et le développement de la bio-informatique en sont des exemples célèbres.

Les prix Nobel de physique et chimie 2024 ont mis en évidence l'apport de l'IA à la découverte scientifique.

Depuis toujours, le CNRS porte l'ambition de contribuer à positionner la recherche française au meilleur niveau international dans cette dynamique, par le recrutement de scientifiques à la pointe, par le soutien aux équipes de recherche, par la structuration des communautés scientifiques, la mise à disposition de moyens de calcul et aussi par des coopérations fortes avec le tissu économique, ainsi qu'avec des acteurs internationaux majeurs.

Cet intérêt particulier se retrouve dans le contrat d'objectifs et de moyens 2019-2023 du CNRS où l'intelligence artificielle est un des six grands défis sociaux (aux côtés du changement climatique, de la transition énergétique, de santé et environnement, des territoires du futur et ses inégalités éducatives) auxquels le CNRS se donnait l'ambition de contribuer de manière particulièrement active et volontariste.

En 2021, le CNRS a ainsi créé le centre AI for science, science for AI (AISSAI) pour faire dialoguer toutes les disciplines scientifiques autour de l'IA et faire émerger des projets interdisciplinaires. Le centre AISSAI a inspiré la déclaration des 6 principaux organismes de recherche européens sur l'IA.

Cette stratégie se déploie également à l'international. Le CNRS a des bureaux et des laboratoires de recherche sur tous les continents, dont des partenariats en IA très visibles (au Canada, au Japon et à Singapour, notamment) ou en développement (en Australie et au Maroc).

Et dans le dernier contrat d'objectifs, de moyens et de performance, 2024-2028, du CNRS c'est « L'IA générative pour les sciences » qui est mise en avant aux côtés du cerveau, de la vie dans l'univers, des matériaux du futur, de l'instrumentation sans limite et des sociétés en transitions.

Une priorité, que s'est donné le CNRS pour accompagner cette dynamique générale, est de mettre à disposition de l'écosystème de recherche et d'innovation un supercalculateur convergé HPC/IA souverain, en partenariat avec la société civile GENCI, avec le soutien constant du Ministère en charge de la recherche (MESR) et du secrétariat général

pour l'investissement (SGPI) dans le cadre de la stratégie nationale en intelligence artificielle (SNIA).

Le supercalculateur Jean Zay n'a cessé d'évoluer pour suivre la demande des utilisateurs. La 4^e extension installée en 2024 lui permet de faire 126 millions de milliards d'opérations par seconde !

Le succès de Jean Zay tient pour beaucoup aux conditions offertes aux utilisateurs : un accès rapide et simplifié pour les « petits » projets et un réseau d'ingénieur(e)s qui assure le support et la formation des équipes utilisatrices.

Ce modèle doit être encore renforcé et ouvert à l'Europe dans le cadre du projet AI Factory France porté par GENCI. En outre, des partenariats avec des acteurs de l'industrie cloud européenne doivent permettre de faire émerger un écosystème d'IA décentralisé, interopérable et souverain, capable de rivaliser à l'échelle mondiale.

Au-delà du calcul et du support, le CNRS développe une stratégie intégrée pour accélérer la découverte scientifique grâce à l'IA, dont un des volets essentiels est le développement des talents. Le CNRS a créé, en 2021, des postes spécifiques pour recruter des scientifiques d'autres disciplines, acculturés à la science des données et à l'IA. Le centre AISSAI lance des appels à postdoctorants pour favoriser le développement de compétences hybrides. Enfin, le CNRS porte le programme Choose France - Rising Talents de la SNIA, ouvert à toutes les disciplines.

Signe incontestable du succès de cette politique de long terme, plus d'une start-up sur cinq issues chaque année des unités de recherche conjointes au CNRS et à ses partenaires ont une composante significative d'intelligence artificielle dans leurs produits ou procédés, en quasi-totalité en lien avec une ou plusieurs autres disciplines. C'est probablement à ces intersections, X+IA (X pouvant être médicaments, matériaux, satellites, vêtements connectés...), que se situent les perspectives d'innovation les plus prometteuses en France et en Europe.

Cependant, le déploiement de l'IA dans les sciences, comme dans la société, n'est pas sans danger. Les équipes de recherche sont mobilisées pour analyser les risques, les contrôler et développer des approches alternatives.

Le coût énergétique et l'empreinte carbone de l'IA représentent un handicap majeur. À court terme, il s'agit d'optimiser les algorithmes et les infrastructures

« bas carbone » tout en développant les bases d'une IA frugale (notamment en privilégiant des modèles avec moins de paramètres) et de mettre en place des politiques de modération de l'usage. À plus long terme, les recherches sur les nouveaux composants ou paradigmes de calcul (quantique, sur ADN, voire biomorphique) pourraient ouvrir des voies alternatives.

Une autre priorité est de développer une IA plus sûre. Le développement de l'IA pour les sciences suppose de pouvoir vérifier, expliquer les résultats produits ou assistés par l'IA. Or les biais présents dans les données scientifiques peuvent se propager et s'amplifier, l'interprétation des résultats est rendue difficile du fait de la complexité croissante des modèles, les systèmes d'IA restent fragiles face à la présence de données erronées ou à des attaques.

L'usage de l'IA dans la pratique de la science soulève également des questions d'intégrité scientifique en lien avec les règles de publication et d'évaluation par les pairs des articles scientifiques, le manque de traçabilité des données et la non-reproductibilité des résultats.

Le développement de l'IA pour les sciences soulève enfin des enjeux de société : inégalités d'accès pour les communautés scientifiques liées au coût et à la concentration des moyens de calcul et des données chez quelques acteurs, dépendance accrue aux outils propriétaires, risque d'erreurs massives ou de manipulation de résultats scientifiques.

Tout cela nécessite de renforcer la transparence, l'explicabilité et la gouvernance de l'IA scientifique. Il faut pour cela une mobilisation de la communauté scientifique à l'échelle européenne, a minima.

L'histoire ne s'arrête pas plus qu'elle ne commence avec ChatGPT. Les équipes de recherche, pleinement conscientes à la fois des potentialités et des limites que présente l'IA générative, travaillent au développement de modèles plus fiables, frugaux, puissants qui permettront, demain, d'avoir des IA agentives capables de prise de décision et d'action autonomes, puis des IA « physiques » pouvant interagir avec le monde physique et les humains.

Il nous appartient de faire en sorte que l'IA devienne un bien commun qui profite à tous et toutes, et ne soit pas accaparé par quelques-uns, États, grandes multinationales ou individus. C'est un enjeu de démocratie planétaire. ■

DU CONTRÔLE-COMMANDE À L'IA GÉNÉRATIVE : 40 ANS DE TRANSFORMATION NUMÉRIQUE DE L'INDUSTRIE

Patrice Caine

Président-directeur général du groupe Thales

Président de Association nationale de la recherche et de la technologie (ANRT)

L' introduction du numérique dans l'industrie est déjà une histoire ancienne. On peut, pour simplifier, la faire débuter dans les années 1980, avec, entre autres exemples, la transformation des systèmes de contrôle-commande et la création des automates programmables ou encore l'arrivée des logiciels de CAO/FAO. À cette époque, l'objectif est d'augmenter la productivité, de stabiliser la qualité, et d'améliorer la répétabilité des processus industriels.

Dans les années 2000, avec la généralisation de l'informatique en réseau et des ERP, le numérique devient un levier de pilotage transversal. On connecte des fonctions organisées jusque-là en silos : production, logistique, maintenance, finance, RH. L'usine devient un système intégré, piloté par la donnée, capable de mieux s'adapter à la demande, d'optimiser ses stocks, de fiabiliser ses plannings.

La décennie 2010 marque une inflexion majeure : c'est l'émergence de ce qu'on a parfois désigné sous l'étiquette d'« Industrie 4.0 ». L'Internet des objets industriels, la robotique, les jumeaux numériques, la réalité augmentée, le cloud ou encore les big datas s'invitent dans les centres d'ingénierie. Les équipements communiquent, apprennent, s'auto-adaptent. L'approche devient résolument systémique, combinant le physique et le virtuel, la simulation et la donnée réelle, l'homme et la machine.

Aujourd'hui, nous abordons une nouvelle phase de la révolution numérique. L'intelligence artificielle générative, les systèmes autonomes, le calcul haute

performance (et dans le futur les ordinateurs quantiques), permettent d'atteindre une complexité de systèmes, une profondeur de modélisation, et une vitesse d'innovation sans précédent.

LE NUMÉRIQUE DÉSORMAIS OMNIPRÉSENT CHEZ THALES

Nous avons eu la chance chez Thales de bénéficier d'une position très favorable pour tirer avantage de ces technologies au fil des décennies. En effet, au-delà d'être un outil au service de l'efficacité de notre production, le numérique se trouve au cœur de nos solutions et de notre R&D. Notre expertise en matière de cybersécurité nous a aussi donné les moyens d'intégrer des innovations à nos unités de production tout en garantissant un très haut niveau de protection et de fiabilité.

Dans nos activités spatiales par exemple, la fabrication d'un satellite repose aujourd'hui sur des jumeaux numériques extrêmement détaillés, répliquant comportements thermiques, contraintes mécaniques, interactions électromagnétiques, etc. Ces modèles, nourris par de grandes quantités de données d'essai ont de multiples avantages : ils nous permettent de réduire les itérations physiques, de reconfigurer rapidement les lignes de production et d'anticiper d'éventuelles défaillances.

À Cannes, le robot SAPHIR, développé pour la pose automatique d'inserts sur les panneaux de satellites, associe vision artificielle, mécatronique de haute précision et analyse de données massives. Cette



combinaison a permis d'accroître la cadence de production tout en renforçant la qualité et la traçabilité. À Hasselt, en Belgique, nous avons inauguré en 2019 une usine de panneaux photovoltaïques conçue nativement selon les principes de l'« Industrie 4.0 ». Elle intègre ainsi des flux de données temps réel et des interfaces numériques intersites, directement connectées à notre centre d'expertise cannois. À la clé : davantage d'agilité et une meilleure gestion des ressources.

Dans le domaine de la défense, le développement de systèmes complexes repose désormais sur des chaînes numériques de bout en bout. Ingénierie système, modélisation comportementale, simulation opérationnelle, tests virtuels, cybersécurité : tout est intégré dès la phase amont.

Le numérique a aussi permis de faciliter le travail en commun entre les équipes de Thales et notre écosystème de PME, d'ETI et de laboratoires partenaires, grâce à des environnements collaboratifs sécurisés.

UNE MUTATION PAS ENCORE ACHEVÉE

Cette numérisation profonde de la chaîne industrielle a transformé les métiers et les compétences. Elle modifie également la place de l'usine dans l'écosystème global. Cette dernière a vocation à devenir un nœud d'un vaste

système industriel distribué, capable d'interagir dynamiquement avec ses parties prenantes. Mais parvenir à ce stade de développement implique au préalable des investissements massifs dans les infrastructures numériques, la cybersécurité, la formation.

L'industrie française avance résolument dans ce sens, et dans tous les secteurs. Mais notre pays doit encore accélérer, notamment pour faciliter l'appropriation des technologies numériques par les PME ou encore pour mettre en place de plateformes numériques mutualisées à l'échelle des filières.

C'est un sujet dont on ne saurait trop souligner l'importance, car le numérique est dorénavant beaucoup plus qu'un outil parmi d'autre. Il constitue véritablement la colonne vertébrale de l'industrie du XXI^e siècle. Il a d'ores et déjà transformé nos manières de concevoir, de produire, de coopérer. Demain, il sera la pierre angulaire de la création d'une industrie plus résiliente, plus souveraine et plus durable. ■





TECHNOLOGIES



INTRODUCTION : TECHNOLOGIES, DES MUTATIONS PROFONDES

Jean-François Prevéraud
Journaliste

Les technologies numériques, toujours plus présentes dans nos vies professionnelles et personnelles, évoluent constamment, tant au niveau des composants, des physiques mises en œuvre, que des architectures systèmes et des logiciels associés. Elles ne suivent plus la 2^e Loi de Moore.

PERFORMANCE

Développer de nouvelles technologies de puces devient prohibitif, les fondeurs doivent s'allier pour mutualiser les coûts faramineux de nouvelles usines, les ingénieurs repenser les processeurs. Exit les gros chips monolithiques, vive les chipelets qui, sur un même substrat, associent des éléments de nature et de technologies différentes, pour obtenir les performances attendues à moindre coût, et garantir une interconnexion à haute densité, garante d'une bande passante extrêmement élevée indispensable au HPC.

Les CPU se voient complétés au sein d'une architecture système hybride par une multitude d'accélérateurs. Par-delà les GPU, ils sont aujourd'hui épaulés par des Neuronal Processing Unit (NPU), des Data Processing Unit (DPU), des IPU (Infrastructure Processing Unit) ... qui les déchargeant de multiples tâches. On voit aussi apparaître de nouvelles physiques, le quantique, la photonique, le spintronique, le neuromorphique... qui changeront la donne dès qu'elles seront apprivoisées. L'informatique classique cédera alors le pas à une informatique hybride, mixant les technologies en fonction des besoins, ce qui ne sera pas sans impact sur les architectures mises en œuvre. La conteneurisation des applications en est un exemple.

FRUGALITÉ

Mais cela ne doit pas faire oublier que le numérique est l'un des plus gros consommateurs d'énergie, 11 % de la consommation électrique en France d'après l'Ademe, ce qui impose aux ingénieurs de développer des puces performantes beaucoup plus frugales. C'est d'autant plus nécessaire que l'IA Générative fait exploser la demande de puissance.

Des évolutions technologiques de composants et d'architectures système qui ont aussi un impact sur le stockage de données. Rien ne sert d'avoir des calculs rapides, s'il faut attendre après la donnée. La performance globale d'un système est celle du maillon le moins performant. C'est pourquoi les fabricants font évoluer leurs substrats et leurs technologies d'accès, d'autant plus que le stockage se fait de plus en plus dans le cloud. Mais on commence déjà à parler de Edge-Storage.

SÉCURITÉ

Enfin, il ne faut pas oublier la sécurité de tels systèmes. Dans un monde en crise où les alliances se renversent, où des guerres éclatent, où les entreprises s'épient, il faut être de plus en plus vigilant face aux cybermenaces pouvant venir de toute part. La donnée et son intelligence sont aujourd'hui la vraie valeur ajoutée des organisations, de l'entreprise aux États. Il faut la protéger.

Autant d'évolutions présentes et à venir que nous avons demandé aux meilleurs experts internationaux d'explicitier dans les pages qui suivent.

EVOLUTION DES COMPOSANTS POUR LE CALCUL HAUTE PERFORMANCE : PERSPECTIVES À 5-10 ANS

Marc Duranton
Senior Fellow, CEA

Les systèmes HPC ont vu une augmentation incomparable de leurs performances en 20 ans : entre le BlueGene/L de 2005 avec ses 280 TFlops et El Capitan de 2024 (1742 Pflops) soit une augmentation de la puissance de calcul de 6222 pour une augmentation modérée de la consommation, de 1,43 MW à 29,6 MW soit globalement une amélioration de l'efficacité énergétique d'un facteur 300.

Cependant, on se trouve à un carrefour : l'essor des charges de travail émergentes — notamment l'intelligence artificielle — et le ralentissement du « scaling » CMOS classique redéfinissent les composants de base. Au cours des dix prochaines années, les systèmes HPC reposeront de plus en plus sur des architectures

spécialisées et hétérogènes pour maximiser les performances dans des contraintes strictes de puissance et de coût : systèmes à base de chiplets et d'accélérateurs dédiés jusqu'aux dispositifs utilisant de nouvelles technologies, comme la photonique, la spintronique et de nouvelles technologies de mémoires et de communication qui devront relever les défis de la consommation électrique, des limites thermiques et de la fiabilité.

ARCHITECTURES À BASE DE CHIPLETS ET PACKAGING AVANCÉ

Au cours des prochaines années, les processeurs et circuits pour le HPC adopteront de plus en plus des conceptions fondées sur des chiplets plutôt que sur



Figure 1 - Augmentation de l'efficacité énergétique des machines HPC en 20 ans

EVOLUTION DES COMPOSANTS POUR LE CALCUL HAUTE PERFORMANCE : PERSPECTIVES À 5-10 ANS

des puces monolithiques. Les chiplets sont de petits modules de silicium, chacun réalisant une fonction, et connectés sur un substrat silicium qui permet une interconnexion à haute densité entre eux : l'interposeur. L'ensemble est ensuite intégré dans un boîtier comme pour un circuit monolithique. Cette approche permet d'associer dans un seul module – « circuit » – des pièces fabriquées sur des nœuds technologiques différents et spécialisées pour des tâches distinctes. Ainsi, un chiplet de cœur CPU haute performance gravé en 5 nm peut être couplé à des chiplets analogiques ou d'entrées/sorties produits sur des nœuds plus anciens et moins coûteux. Le moteur de cette évolution est à la fois économique et pratique : la fabrication d'une seule grosse puce devient prohibitive et se caractérise par de faibles rendements, alors que plusieurs puces plus petites améliorent le rendement et réduisent le coût par composant. Elle permet aussi d'intégrer côte à côte une large palette d'accélérateurs et de mémoires et de garder de la diversité et de la spécialisation sans le coût de refaire entièrement une puce monolithique. Des innovations de packaging avancé soutiennent cette révolution des chiplets. Les interposeurs silicium avec TSV (Through Silicon Via) (intégration 2,5D) et l'empilement direct de dies (3D) offrent l'interconnexion dense nécessaire pour relier les chiplets avec une bande passante extrêmement élevée. Par exemple, les puces MI350 d'AMD fusionnent chiplets CPU, chiplets GPU et 8 piles de mémoire HBM3 dans une configuration 3D au sein d'un même package pour un total de plus de 150 milliards de transistors. En intégrant CPU et GPU sur un interposeur partagé avec la mémoire, le MI325 obtient une bande passante très élevée, idéale pour les charges IA et HPC.

Malgré leur promesse, les architectures à chiplets doivent résoudre plusieurs défis. L'un des plus urgents est l'absence de normes industrielles pour les interfaces, la distribution d'énergie et les spécifications physiques. L'industrie s'efforce donc de converger vers des standards ouverts ; l'Universal Chiplet Interconnect Express (UCIe) en est un exemple, visant des débits de 32 Gbps par broche entre chiplets. Les tests Known Good Die (KGD) deviennent cruciaux pour éviter qu'un chiplet défectueux n'entraîne la mise au rebut d'un boîtier complet coûteux. Des capteurs embarqués (température, courant, détection d'erreurs) et des mécanismes de redondance sont aussi explorés pour détecter et gérer les défaillances *in situ*.

ACCÉLÉRATEURS IA ET ACCÉLÉRATEURS SPÉCIFIQUES AU DOMAINE

L'essor de l'apprentissage automatique pousse les systèmes HPC à intégrer de nouveaux accélérateurs spécialisés. Depuis une décennie, les GPU dominent grâce à leur parallélisme massif. Désormais, les Unités de Traitement Neuronales (NPU) – processeurs optimisés pour les réseaux neuronaux – gagnent du terrain pour l'inférence à faible latence. Les CPU modernes intègrent aussi des instructions IA ou de petits accélérateurs sur puce. Cette tendance est renforcée par la réduction de la précision numérique : là où les simulations utilisaient la double précision (FP64), les modèles IA s'accommodent souvent de formats 16 bits, 8 bits, voire 4 bits, avec un impact minimal sur la précision, d'où des gains majeurs en performance et énergie (l'arithmétique FP8 peut offrir ~8 × plus de performances que FP64).

D'autres accélérateurs spécifiques émergent. Le matériel reconfigurable comme les FPGA peut accélérer des algorithmes spécifiques ou servir de banc d'essai pour de nouvelles architectures. Les DPU (Data Processing Units) ou IPU (Infrastructure Processing Units) soulagent le CPU des tâches réseau, stockage et sécurité. On voit aussi apparaître des accélérateurs fondés sur l'ISA ouverte RISC-V, illustrant la recherche d'une personnalisation poussée avec moins de contraintes. À mesure que les gains de la technologie CMOS se réduisent, la prochaine décennie verra une profusion d'accélérateurs, chacun ciblant une classe de charges (IA, graphes, cryptographie, mouvements de données) au sein de plates-formes hétérogènes.

CALCUL HÉTÉROGÈNE ET HIÉRARCHIES MÉMOIRE

Avec la multiplication des accélérateurs, l'architecture des systèmes HPC devient un modèle hétérogène où divers processeurs doivent coopérer. Des standards tels que Compute Express Link (CXL) permettront aux CPU, GPU, NPU et dispositifs mémoire d'accéder à un espace mémoire cohérent commun.

La technologie mémoire se diversifie mais le gap entre les capacités de calcul et de mémorisation s'accroît au cours du temps : la HBM (High Bandwidth Memory) fournit un débit énorme mais reste coûteuse et limitée en capacité. Des alternatives comme la DDR5/

SCALING OF PEAK HARDWARE FLOPS, AND MEMORY/INTERCONNECT BANDWIDTH

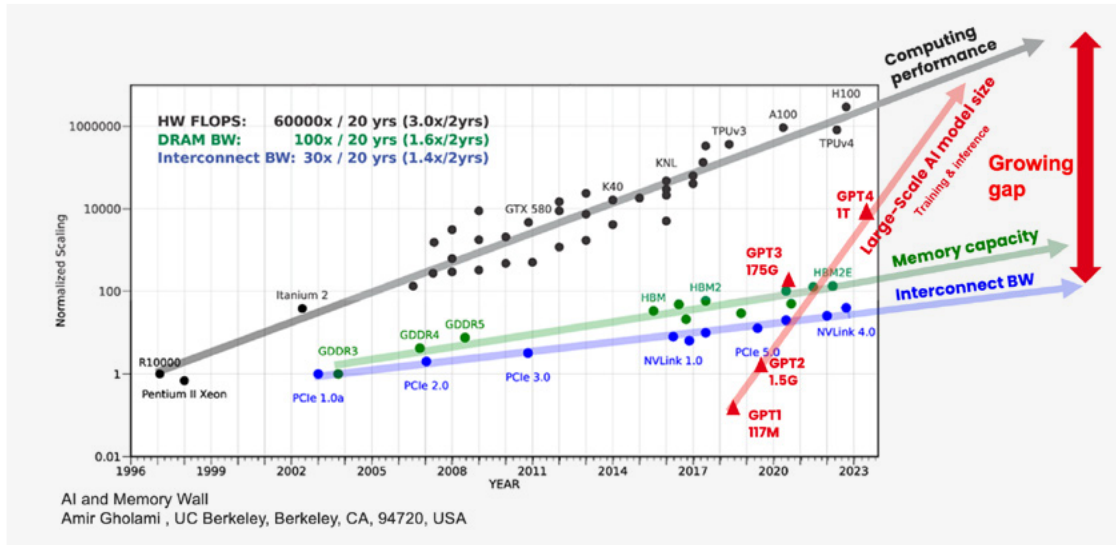


Figure 2 - Augmentation du gap entre les besoins de l'IA, les performances de calcul et des mémoires

LPDDR5X et la MCR-DIMM (Multiplexer Combined Ranks DIMM) promettent un meilleur compromis coût/puissance/bande passante. Des architectures à mémoire hiérarchisée combineront HBM ou SRAM sur boîtier, un grand pool DDR pour la capacité, et potentiellement de la mémoire non volatile ou distante via CXL. Les concepts de processing-in-memory (PIM), encore limités aujourd'hui, pourraient à terme réduire la circulation des données en exécutant certaines opérations au sein même des puces mémoire.

Les interconnexions spécialisées (NVLink, Infinity Fabric, etc.) relient des grappes de GPU pour former des « super-nœuds » partageant la mémoire, comme le demandent les entraînements de modèles comptant des centaines de milliards de paramètres sur des milliers de GPU. L'avenir verra sans doute des architectures désagrégées où les ressources (mémoire, accélérateurs) sont allouées dynamiquement sur un tissu réseau, soutenues par des liaisons optiques et des protocoles comme CXL. L'objectif reste de minimiser la latence et le coût énergétique du mouvement de données. Une architecture HPC hétérogène réussie rendra la présence de multiples puces spécialisées aussi transparente que possible pour les

programmeurs, en atteignant la performance et l'efficacité comme s'il s'agissait d'une seule machine unifiée.

TECHNOLOGIES ÉMERGENTES : PROCESSEURS NEUROMORPHIQUES, PHOTONIQUES ET SPINTRONIQUES

La photonique gagne du terrain pour la communication et même le calcul. Les interconnexions optiques co-packagées (CPO) pourraient remplacer les liens cuivre, apportant des débits immenses et une baisse de l'énergie par bit transmis. Des accélérateurs photoniques capables de multiplications matricielles optiques apparaissent aussi. Les défis concernent l'intégration, l'impact thermique des lasers on-chip et la fiabilité, mais le gain potentiel en bande passante et en efficacité énergétique nourrit un effort de R&D intense. On voit apparaître des interposeurs photoniques

La spintronique exploite le spin de l'électron pour la mémoire et la logique. Les mémoires MRAM non volatiles en sont l'exemple commercial le plus proche. Elles pourraient servir de cache persistant rapide ou de stockage de checkpoint en HPC. À plus long terme, des réseaux de spins (machines d'Ising) pourraient résoudre

EVOLUTION DES COMPOSANTS POUR LE CALCUL HAUTE PERFORMANCE : PERSPECTIVES À 5-10 ANS

des problèmes d'optimisation directement en matériel. Les avancées en matériaux et dispositifs (diminution de la variabilité, passage à l'échelle) détermineront la vitesse d'intégration de la spintronique.

Cela montre la voie aux accélérateurs utilisant des phénomènes physiques pour faire du calcul. Un exemple est illustré par les processeurs neuromorphiques, fonctionnant de façon asynchrone et événementielle, avec une efficacité énergétique extrême pour certains algorithmes. Cependant, des challenges liés au passage à l'échelle restent à résoudre : ces dispositifs ne supportent souvent pas de repliement temporel comme les architectures numériques classiques : il faut donc autant de dispositifs que la taille du problème à résoudre car la mémorisation de leur état et la reconfiguration pour travailler sur une autre partie du problème restent peu efficaces. Ces dispositifs ont souvent une variabilité qui limite aussi le passage à l'échelle. Dans l'horizon 5-10 ans, ces puces resteront des coprocesseurs spécialisés mais pourraient influencer en profondeur la conception de systèmes ultra-parallèles et sobres en énergie.

EFFICACITÉ ÉNERGÉTIQUE, GESTION DE LA PUISSANCE ET THERMIQUE

La consommation énergétique constitue la contrainte ultime : une machine exascale doit rester dans une enveloppe de l'ordre de 20-30 MW. Les accélérateurs spécialisés, la réduction de précision et les optimisations

DVFS (adaptation du point de fonctionnement) visent à maximiser les flops par watt. Augmenter encore l'efficacité exigera des approches disruptives : fonctionnement à tension proche du seuil, nouveaux matériaux, architectures 3D et spécialisation accrue.

ARCHITECTURE SYSTÈME, FIABILITÉ ET DURABILITÉ

La convergence de ces tendances transformera l'architecture des supercalculateurs, moins monolithique et plus composé de ressources variées. Mais cette complexité accroît le risque de défaillances. Des études montrent que les pannes sont déjà fréquentes et augmenteront avec le nombre de composants. Le MTTF (temps moyen avant échec) pourrait être de 14 mn pour un job de 131072 GPU*. Les futurs systèmes intégreront donc des mécanismes de tolérance aux pannes à plusieurs niveaux, des capteurs on-chip aux algorithmes de maintenance prédictive basés sur la télémétrie. Les nouvelles technologies, nécessaires pour continuer l'augmentation des performances et de l'efficacité sont encore peu caractérisées à grande échelle, et nécessiteront des phases d'intégration prudentes pour garantir les contraintes de fiabilité. ■

L'INFORMATIQUE QUANTIQUE : VERS UNE NOUVELLE ÈRE DU CALCUL INTENSIF

Pascale Senellart

Directrice de Recherche CNRS

Centre de Nanosciences et de Nanotechnologies – Université Paris Saclay, CNRS

Co-fondatrice et conseillère scientifique Quandela

La décennie écoulée, et très probablement la suivante, constitue une période charnière pour l'introduction de l'informatique quantique dans le paysage du calcul intensif. L'évolution rapide de ce domaine est remarquable. Il y a une vingtaine d'années, les technologies quantiques dites de « seconde génération » relevaient encore d'une perspective lointaine, confinées aux laboratoires académiques. En dehors de quelques initiatives pionnières telle que celle de D-Wave, le domaine a véritablement connu un tournant en 2016, avec les premières machines de calcul quantique mises en ligne par IBM et l'engagement d'autres grandes entreprises dans cette course technologique. L'écosystème s'est ensuite élargi rapidement, avec la création de la majorité des jeunes entreprises spécialisées dans le calcul quantique entre 2015 et 2020.

Aujourd'hui, les premiers ordinateurs quantiques commencent à être intégrés dans les infrastructures de calcul haute performance (HPC) à l'échelle mondiale. L'Europe, à travers le programme EuroHPC, investit également dans cette trajectoire. Les promesses associées à cette technologie sont considérables : au-delà d'une simple accélération, elle ouvre la voie à des calculs inaccessibles aux supercalculateurs les plus puissants, grâce à l'introduction de nouvelles primitives de calcul, fondées sur les propriétés fondamentales de la physique quantique. De telles perspectives ont conduit de nombreux États à considérer le calcul quantique comme une technologie stratégique et souveraine. Toutefois, les défis à relever demeurent considérables,

qu'il s'agisse de l'intégration dans les usages actuels du HPC, de l'augmentation des performances matérielles, ou encore du développement d'applications pertinentes.

LES FONDEMENTS ET PROMESSES DE LA SECONDE RÉVOLUTION QUANTIQUE

Les architectures actuelles des supercalculateurs tirent déjà parti des concepts de la mécanique quantique, notamment à travers l'invention du transistor ou du laser au milieu du XX^e siècle. Cette « première révolution quantique » a façonné l'ère numérique contemporaine. L'impact de cette étape historique permet de mesurer le potentiel de la « seconde révolution quantique », centrée cette fois sur l'exploitation de phénomènes plus subtils tels que l'intrication et la superposition quantiques.

Le passage du bit classique au bit quantique (qubit), capable de représenter simultanément les états 0 et 1, ouvre la voie à de nouveaux paradigmes algorithmiques. Certains algorithmes théoriques, comme celui de Shor pour la factorisation de grands nombres premiers, illustrent la puissance disruptive de cette approche. D'autres, plus récents, promettent des accélérations exponentielles ou polynomiales dans des domaines tels que la chimie quantique, la conception de nouveaux matériaux, la finance ou la logistique.

Ces perspectives se heurtent cependant à une contrainte majeure : la cohérence quantique. Celle-ci, indispensable au maintien de la superposition et de l'intrication, est extrêmement fragile et tend à disparaître sous l'effet

L'INFORMATIQUE QUANTIQUE : VERS UNE NOUVELLE ÈRE DU CALCUL INTENSIF

des perturbations de l'environnement. Depuis plusieurs décennies, la communauté scientifique s'efforce d'isoler et de contrôler ces systèmes, et les progrès réalisés au cours des dix dernières années ont permis la mise en œuvre des premiers algorithmes sur des dispositifs quantiques. Néanmoins, le contrôle optimal et à grande échelle demeure un objectif de long terme.

DIVERSITÉ TECHNOLOGIQUE ET ENJEUX DU PASSAGE À L'ÉCHELLE

De manière analogue aux débuts de l'informatique classique, les ordinateurs quantiques reposent aujourd'hui sur une pluralité de technologies matérielles. Les qubits peuvent être implémentés au moyen de circuits supraconducteurs, de photons, de composants à base de silicium, ou encore d'atomes uniques piégés dans le vide.

Chaque approche présente des avantages spécifiques mais aussi des limites. Les systèmes photoniques ou atomiques offrent une conservation naturelle de la cohérence. En revanche, pour passer à des échelles de plusieurs millions de qubits, des architectures basées sur des matériaux supraconducteurs ou semi-conducteurs apparaissent mieux adaptées à l'industrialisation, bien que leur stabilité quantique soit plus difficile à préserver. À terme, il est probable que les architectures hybrides, combinant différentes approches, constituent la voie la plus prometteuse pour atteindre des performances universelles.

DU DÉVELOPPEMENT EXPÉRIMENTAL AUX PREMIERS CAS D'USAGE

Le développement initial du calcul quantique a été porté par des physiciens et, progressivement, par des spécialistes en algorithmique quantique. Les interactions avec les communautés du HPC et de l'informatique appliquée, bien que tardives, sont désormais en forte croissance. Cette convergence est indispensable pour accélérer l'identification de cas d'usage pertinents.

Les premières expérimentations reposent sur des dispositifs encore limités et imparfaits. Cependant, l'accessibilité en ligne de processeurs quantiques ainsi que leur intégration dans plusieurs centres de calcul en Europe, aux États-Unis et en Asie, ont permis la multiplication d'études exploratoires. Des acteurs tels qu'IBM, Quandela, Quantinuum ou Pasqal testent

des applications dans des domaines variés : optimisation de processus industriels, simulation chimique, ou encore apprentissage automatique. L'articulation entre intelligence artificielle et calcul quantique représente également une dynamique de recherche croissante. Des initiatives récentes, telles que le défi Perceval Quest¹ ou le Quantum Challenge Airbus-BMW², illustrent cette convergence, où le quantique ouvre des perspectives pour l'IA tandis que l'IA contribue à l'optimisation des architectures quantiques.

Au-delà de la question des qubits et des spécificités matérielles propres à chaque plateforme, l'apport fondamental du calcul quantique réside dans l'introduction de nouvelles primitives de calcul. Là où les architectures classiques n'ont évolué qu'en échelle depuis plusieurs décennies — parallélisme massif, accélérateurs graphiques ou vectoriels — les processeurs quantiques reposent sur des mécanismes totalement différents, issus de la superposition et de l'intrication. Ces primitives non conventionnelles ne remplacent pas les ressources de calcul intensif existantes : elles les complètent. Le défi majeur consiste à intégrer ces briques inédites au sein des chaînes de calcul haute performance, en tirant parti des supercalculateurs pour la majorité des tâches tout en déléguant certaines étapes vers les processeurs quantiques. Les perspectives applicatives sont doubles : d'une part, dans le domaine de l'apprentissage automatique quantique où l'expressivité des modèles peut être enrichie par des circuits quantiques ; d'autre part, dans des algorithmes spécialisés exploitant ces primitives (recherche combinatoire, simulation de systèmes quantiques complexes, optimisation). Dans ce contexte, il ne s'agit pas de comparer directement les performances des machines quantiques et classiques, mais de concevoir de nouveaux flux hybrides de calcul, capables de combiner la robustesse des infrastructures HPC avec les capacités émergentes des processeurs quantiques.

À court terme, l'objectif est d'atteindre le régime dit « d'utilité quantique », caractérisé par des gains mesurables en temps de calcul, en consommation énergétique ou en précision, grâce à l'intégration conjointe de processeurs classiques et quantiques. Cette étape transitoire favorisera la maturation des technologies et guidera les développements futurs en vue d'un calcul quantique universel avec correction d'erreurs.

1. <https://www.scaleway.com/en/blog/1st-perceval-quest-a-journey-into-quantum-machine-learning/>

2. <https://www.airbus.com/en/innovation/digital-transformation/quantum-technologies/airbus-and-bmw-quantum-computing-challenge>

DÉFIS, RIGUEUR ET PERSPECTIVES

La réussite de cette transformation repose sur la structuration d'un écosystème robuste. Il s'agit notamment de développer de nouveaux paradigmes de programmation, de rendre accessibles les outils de développement quantique, et de favoriser l'élargissement de la communauté d'utilisateurs.

Deux écueils doivent être évités. D'une part, un excès de scepticisme, qui sous-estime les progrès réalisés et les perspectives à moyen terme; d'autre part, un excès de communication, parfois trompeur, qui contribue à entretenir des attentes irréalistes. Entre ces deux extrêmes, le chemin à suivre doit être marqué par la rigueur scientifique et l'ambition technologique.

Le calcul quantique constitue en effet une aventure scientifique singulière, mobilisant des compétences pluridisciplinaires et suscitant un engouement comparable à celui des grandes révolutions technologiques du siècle précédent. Dans une dizaine d'années, il sera possible de mesurer le chemin parcouru, et de situer la place qu'aura prise cette technologie dans l'écosystème du calcul intensif. Si son avenir ne se résume pas à un « calcul plus rapide », mais bien à l'intégration de nouvelles primitives non conventionnelles au sein des infrastructures HPC, son impact sera profond. Tout indique que l'informatique quantique s'imposera comme l'une des ruptures majeures du XXI^e siècle. ■

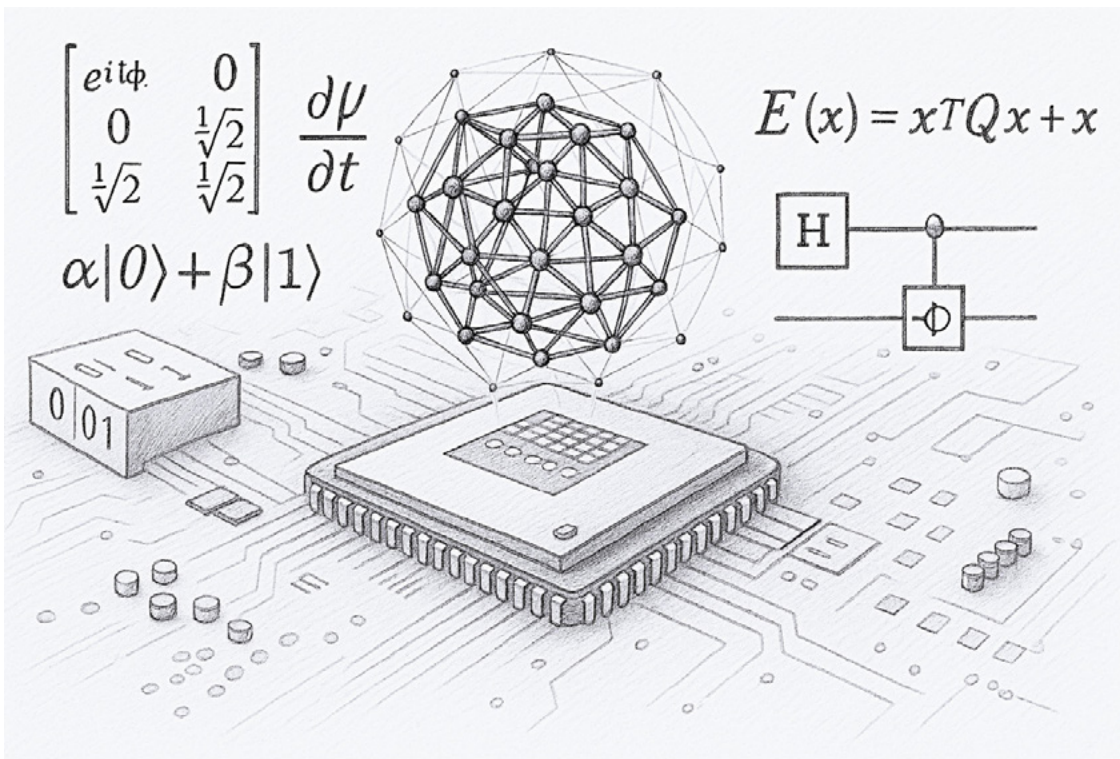


Illustration générée par IA

LES DÉFIS ÉNERGÉTIQUES DES ORDINATEURS QUANTIQUES

Olivier Ezratty

Quantum Energy Initiative, cofondateur

L'une des motivations de la création de la Quantum Energy Initiative (QEI) en 2022 était de fédérer les chercheurs académiques et les industriels s'intéressant à la question de la consommation énergétique des technologies quantiques, et en particulier des ordinateurs quantiques. Il s'agissait d'étudier la question le plus en amont possible de la conception de ces systèmes pour éviter les mauvaises surprises qui abondent dans l'intelligence artificielle avec des projets pharaoniques de centres de calcul consommant des GW de puissance électrique.

L'approche consiste à créer les bases scientifiques de la compréhension, de l'évaluation et de l'optimisation de cette consommation en partant de la physique fondamentale des qubits pour intégrer ensuite l'ensemble des autres composantes des ordinateurs quantiques dont l'électronique de contrôle des qubits, la cryogénie, ainsi que toutes les piles logicielles. L'une des actions clés de cette QEI est l'animation d'un groupe de travail à l'IEEE pour standardiser la manière de mesurer la consommation d'énergie des ordinateurs quantiques.

Depuis 2022, l'écosystème du calcul quantique a évolué avec la multiplication de démonstrations de qubits logiques, notamment supraconducteurs, à atomes froids ou ions piégés, et la publication de feuilles de route détaillées par les constructeurs qui sont mobilisés dans la création d'ordinateurs à tolérance aux fautes. La correction d'erreurs nécessaire à la création de ces machines présente un coût élevé. Elle multiplie par un à plusieurs ordres de grandeur les infrastructures matérielles du

calcul quantique par rapport à celles des calculateurs bruités actuels dits NISQ (Noisy Intermediate Scale Quantum). La question se pose donc de l'intégration de ces futurs ordinateurs quantiques dans les centres de calcul, notamment ceux qui intègrent déjà des supercalculateurs et l'expérience associée. Comment les infrastructures de calcul classiques devront-elles s'adapter à l'intégration de ces nouveaux venus ?

Cela soulève un grand nombre d'interrogations. La première concerne l'usage qui sera fait de ces machines. Seront-elles dédiées à des calculs très spécialisés relevant de la recherche fondamentale et appliquée, comme pour la réalisation de simulations chimiques ou de nouveaux matériaux, ou étendues à des applications de production utilisées au quotidien par un grand nombre d'entreprises, par exemple pour résoudre des problèmes d'optimisation ou d'apprentissage automatique ? Quelles en seront la valeur ajoutée scientifique et économique pour les utilisateurs, qu'ils soient issus des mondes académiques ou industriels ?

ESTIMER LA PUISSANCE CONSOMMÉE PAR LES ORDINATEURS QUANTIQUES

Les premières estimations issues des fournisseurs indiquent que la puissance électrique nécessaire à l'alimentation d'ordinateurs quantiques supportant quelques milliers de qubits logiques pourrait aller d'un rapport de 1 à 100 selon les technologies, avec un écart similaire en termes de coût des machines et d'empreinte carbone. Soit, de quelques dizaines à centaines de kW

à une centaine de MW. Dans certains cas, ces machines seront donc plus coûteuses en puissance que les plus grands supercalculateurs actuels qui vont de 1 MW à 40 MW à pleine charge. Cela correspondra à des situations où ces machines pourront résoudre des problèmes inaccessibles aux supercalculateurs classiques.

Se posera alors la question de leur acceptabilité énergétique. Elle dépendra de leur retour sur investissement, des pratiques actuelles et à venir d'investissements en calcul haute performance et des capacités et contraintes des réseaux électriques.

Tout ceci s'accompagne d'autres contraintes opérationnelles liées entre autres à la cryogénie, nécessaire pour quasiment tous les types de qubits, la stabilité en température et en vibrations et la maintenance de certains composants.

INTÉGRER LES COÛTS CLASSIQUES DU CALCUL QUANTIQUE

Une dernière contrainte énergétique rarement évaluée correspond aux coûts du calcul classique qui environne les processeurs quantiques. Ceux-ci correspondent à différentes fonctions clés pour lesquelles les estimations se font encore rares.

La première correspond au coût classique engendré par la correction d'erreurs en temps réel. Avec un grand nombre de qubits physiques supportant les qubits logiques corrigés nécessaires au calcul à tolérance aux fautes, cette correction repose sur de nombreux cycles de lecture et de détection de syndromes. C'est une opération coûteuse en électronique de contrôle, en traitement du signal puis, dans certains cas, en calcul classique.

La seconde est celle de la compilation et de l'optimisation des circuits quantiques. Avec des milliers à des millions de qubits physiques, cette compilation pourrait présenter des temps de calcul et coûts énergétiques significatifs, surtout lorsque l'opération est renouvelée fréquemment au gré de l'évolution des données du

problème. Elle devra aussi intégrer les techniques de partitionnement de circuits quantiques nécessaires à la répartition du calcul quantique sur plusieurs processeurs quantiques reliés entre eux par des moyens d'interconnexion quantiques.

La troisième correspond au coût du pré et du post-traitement classique des algorithmes quantiques. Dans les simulations chimiques, ces traitements sont conséquents. Il faut notamment préparer le circuit décrivant l'état quantique du système quantique à simuler, qui doit être aussi proche que possible de l'état fondamental de ce système.

Ces trois tâches classiques ont en commun de devoir parfois résoudre des problèmes « NP complets », à savoir, que leur coût peut augmenter exponentiellement avec l'augmentation de la puissance des ordinateurs quantiques. Mais des optimisations sont toujours possibles. On sera donc amené à devoir évaluer non seulement la consommation d'énergie des processeurs quantiques et de leur environnement immédiat, mais également celle de l'ensemble des outils de calcul classiques nécessaires à leur fonctionnement.

ÉQUILIBRER TEMPS DE CALCUL, PUISSANCE ET ÉNERGIE CONSOMMÉES

Il faudra aussi trouver les bons compromis entre le temps de calcul, le coût et la consommation d'énergie des processeurs quantiques. En effet, certaines estimations chiffrent à plusieurs mois et au-delà la durée de simulations quantiques intéressantes. Pour accélérer ces temps de calcul, on pourra éventuellement répartir l'exécution des circuits quantiques (identiques) sur plusieurs ordinateurs quantiques fonctionnant en parallèle, à condition que leur coût soit acceptable.

Tout ceci confirme le besoin d'adopter une approche systémique et interdisciplinaire intégrant la physique quantique, les technologies habilitantes, l'algorithmie quantique, la correction d'erreurs, le calcul classique ainsi que la dimension économique. ■

ÉVOLUTION DES ARCHITECTURES DE SUPERCALCULATEURS

Jean-Philippe Nominé

Fellow, CEA DIF

Patrick Carribault

Fellow, Directeur de Recherche, CEA DIF

Jean-Christophe Weill

Fellow, CEA DIF

Le calcul intensif a eu recours à des ordinateurs « mainframe » dans les années 1960-1970, puis à des calculateurs dits vectoriels jusqu'à la fin des années 1990 – marché réduit à un nombre limité de grands organismes achetant des machines Cray ou NEC, typiquement. Ces machines très spécialisées présentaient des capacités de mise en pipeline des séquences de traitement de données (organisées en « vecteurs ») et de fortes bandes passantes mémoire. Cette filière s'est progressivement éteinte face à l'essor de machines massivement parallèles à la fin des années 1990. Ces architectures de type cluster ont bénéficié de l'intégration de processeurs de grande diffusion (composants COTS) bien plus économiques que les processeurs vectoriels, et utilisables en masse pour augmenter les performances. Ce modèle prévaut depuis lors, avec une organisation générale de machine en grappe de nœuds plus ou moins complexes et homogènes, partageant de façon plus ou moins uniforme une mémoire locale au nœud. Les nœuds sont interconnectés par un réseau à la performance adaptée, et interconnectés en général par le même réseau à un système de stockage de masse.

MODÈLE D'ARCHITECTURE CLUSTER

Depuis plus de 20 ans, le modèle d'architecture cluster a donc dominé, avec des raffinements et des variantes et évolutions diverses :

- Apparition des multi-coeurs pour continuer à offrir des gains en puissance de calcul sans faire exploser la consommation énergétique ;

- Pour la même raison, plafonnement des cadences d'horloge (1 à 3 GHz) ;
- Ajout d'extensions vectorielles aux CPU ;
- Ajout de GPU en co-processeurs, jusqu'à devenir les composantes principales de la puissance de calcul ; en effet la densité puissance calcul/puissance consommée électriquement est favorable aux GPU, bien que plus difficiles à programmer que le CPU pour la plupart des applications (voire pas exploitables par certaines, du moins pas à un coût raisonnable de portage).

IMPACT DE L'ARCHITECTURE SUR LA PROGRAMMATION

Le parallélisme est de mise à tous les niveaux dans ces approches :

- intra-nœud voire intra-processeur multi-coeurs, par exemple exprimé en OpenMP ou par *threads* explicites ;
- inter-nœuds via l'interconnexion et à base de système de communication par messages, ou avec programmation de tâches réparties (couche de type MPI) ; la mémoire globale du calculateur est l'agrégation de toutes les mémoires locales des nœuds ;
- les entrées-sorties elles-mêmes utilisent des procédés parallèles pouvant regrouper des flux de données partiels sur des serveurs de proximité ou des fichiers partagés ;
- le système de stockage utilise en général un ensemble de serveurs dédiés et effectue des opérations en parallèle (exemple : LUSTRE comme *parallel filesystem*).



Supercalculateur EXA-HE, CEA/DAM – 180 Pflop/s
 BullSequana XH3000, avec processeurs Grace Hopper Superchip 72C 3GHz, NVIDIA GH200 Superchip
 BullSequana eXascale Interconnect (Quad-Rail BXI v2)
 Refroidissement liquide à l'eau tiède

Les principales difficultés d'utilisation efficace de telles architectures sont :

- La nécessaire cohabitation et imbrication de différents niveaux de parallélisme (depuis *l'instruction level parallelism* au niveau des cœurs de calcul, le possible usage d'unités d'extensions vectorielles locales à chaque processeur, les *threads* intra-processeur ou intra-nœud, la distribution de processus ou tâches coordonnées entre nœuds) ;
- Les limitations, et la croissance relative plus lente en performance, des bandes passantes processeur/mémoire ; une instruction de calcul peut dans les cas défavorables nécessiter plusieurs dizaines ou centaines de cycles pour amener les données en registre ou en cache de proximité (on constate qu'une majorité

d'applications est généralement *memory-bound*, c'est-à-dire limitée en performances par cet effet, ne saturant jamais le processeur en calculs) ;

- De manière générale les mouvements de données à tous les niveaux sont une source majeure de dégradation de performance, entre processus d'une simulation, entre simulation et fichiers d'entrées ou de sorties.

D'une certaine manière le HPC est une lutte permanente (sisyphéenne...) pour adapter et optimiser les logiciels sur les architectures qui évoluent toujours plus vite que le matériel. On n'atteint en général pas d'optimum avant de changer de matériel. On sait que bien souvent plus d'efforts sur l'efficacité du logiciel produiraient des gains conséquents (calculs moins

énergivores, plus rapides ; plus de calculs possibles sur les ressources et avec l'énergie disponible) mais les ressources humaines sont hélas souvent la ressource *in fine* la plus rare, et permettent rarement de se rapprocher d'un optimum en la matière.

Cette hétérogénéité est donc devenue dominante, sans que ne se dégage de grand standard de programmation productive et aux performances portables. MPI+X (X=OpenACC, OpenMP par exemple) règne encore sur le HPC. Des méthodes et outils pour mieux abstraire et séparer l'expression du parallélisme de l'expression du modèle à résoudre (plateformes comme Kokkos, DSL...), et d'autres pour traduire source à source et optimiser du code, sont apparus, là encore sans grande voie dominante.

ÉVOLUTION DES CAS D'USAGES : ÉMERGENCE DE L'INTELLIGENCE ARTIFICIELLE

L'intelligence artificielle (Deep Learning, avec ses besoins en inférence et en entraînement) s'est invitée dans le jeu depuis quelques années. Friande de grosse puissance de calcul et traitement, elle a piloté les évolutions du domaine, pas toujours au bénéfice du HPC qui lui avait apporté un socle de techniques à base de parallélisme :

- Les besoins en précision réduite de l'IA détournent les fabricants de processeurs du calcul flottant double précision nécessaire au calcul HPC « haute-fidélité » ; de plus le prix des GPU s'envole, durablement on ne sait pas, mais le HPC notamment public en pâtit pour ses investissements.

Doit-on espérer des fabricants la fourniture d'émulation 64 bits sur des unités arithmétiques 4, 8, 16 bits ? Pas certain. Doit-on adapter les applications autant que faire se peut en précision, et réduire au strict nécessaire la double précision flottante (mixer différentes précisions) ? Vaste problème, peut-être pire que ce que l'on a connu pour paralléliser les anciennes applications vectorielles. Y a-t-il des méthodes bas niveau générales d'émulation logicielle efficaces de précision plus élevée ? Pas encore (cf. Ozaki qui ne fait que des produits de matrices*).

- Les architectures de nœuds tendent vers plusieurs GPU par CPU, et se pose la question de partage mémoire local efficace. La voie un temps entrevue des APU (accelerated processing unit) avec intégration mémoire entre CPU et GPU ne devient pas majoritaire ni certaine.
- Les besoins réseau sont aussi poussés par l'IA, pour l'injection efficace de données massives d'entraînement dans les GPU. Ici on s'éloigne moins du HPC, la cause peut être commune, et par exemple motiver des usages vertueux de la photonique d'interconnexion en HPC et IA.
- L'IA n'a pas les mêmes besoins de gestion des données que le HPC, plus structuré et hiérarchisé en général. Mais pour l'exascale le HPC regarde de plus en plus le stockage objet, pour dépasser des limitations des systèmes de fichiers parallèles actuels, de POSIX.

Entre tensions et divergences, les architectures HPC et IA conserveront un destin lié. Le HPC devra évoluer pour rester dans la course, au service de méthodes et de simulations indispensables pour encore longtemps. ■

*Ozaki : <https://arxiv.org/html/2504.08009v1>

ARCHITECTURES HYBRIDES, MYTHES ET RÉALITÉS

Arnaud Bertrand

Senior Vice President 3DS Outscale R&D

Démarrons en détruisant un premier mythe : « l'Architecture unique », forgée par le Sauron des data centers, qui saurait tout faire parfaitement, n'existe pas.

Tout d'abord une architecture de calcul est faite pour servir un cas d'usage. Elle est conçue avec une idée en tête qui est le meilleur compromis possible pour ce cas d'usage entre :

- Coût ;
- Performance ;
- Durabilité, qui regroupe impact environnemental (incluant l'électricité), opérations (disponibilité) et capacité à évoluer dans le temps.

L'ARCHITECTURE «GENERAL PURPOSE» PAR EXCELLENCE : LE CLOUD COMPUTING

Les architectures de cloud computing ont été imaginées dès l'origine pour fournir un compromis acceptable sur l'ensemble de ces trois axes.

Elles allient :

- Industrialisation des déploiements (choix matériels homogènes, réseau fédérateur, 1-2 types de serveur, stockage partagé) ;
- Allocation des ressources au plus juste au travers d'un orchestrateur ;
- Partage des ressources au travers d'une abstraction, le plus souvent la machine virtuelle (VM).

L'orchestrateur conserve une cartographie en permanence à jour de l'allocation des ressources. Il alloue une quantité de ressources à une abstraction, mais ne maîtrise pas l'usage réel qui va être fait de cette ressource (différence entre nombre de cœurs alloués et taux d'usage processeur réel).

L'orchestrateur permet d'optimiser le coût de possession pour le cloud provider. Par contre, sans contrôle sur l'application cliente, il n'est pas possible d'optimiser l'usage réel des ressources. Il en découle 2 conséquences :

- Des data centers chauffent pour faire tourner des serveurs qui sont occupés à moins de 50 % ;
- Le coût d'achat pour le client n'est pas optimal – cassons aussi le mythe du « le cloud c'est moins cher ». C'est plus pratique, plus rapide à mettre en œuvre, plus simple, mais pas moins cher qu'une architecture optimisée pour un cas d'usage.

LA MACHINE VIRTUELLE, ABSTRACTION REINE DU CLOUD, À NOUVEAU UNE HISTOIRE DE COMPROMIS

L'abstraction utilisée jusqu'à présent était la VM. Elle présente l'avantage d'assurer un excellent cloisonnement de la mémoire et une réservation de ressource fine. Mais a aussi des limites : elle supporte mal les usages « aux limites » (mémoire ou calcul), le mille-feuille applicatif (hyperviseur de virtualisation, émulateur, VM, OS de la VM, middleware) empêche de tirer le maximum des ressources et complexifie le diagnostic.

L'avantage de ce type d'architecture est de répondre aux besoins de calcul de manière convenable dans les cas « moyens », mais de manière médiocre dans les cas « aux limites ».

Pour ces cas-là, il a fallu concevoir des architectures spécialisées :

- Sur la mémoire, au travers des grands systèmes à mémoire cohérente, autrement appelés serveurs « scale-up » ou SMP ;
- Sur le calcul, pour lequel des serveurs optimisés ont été conçus par ex. pour traiter du FP64.

La première a donné lieu à de nombreuses générations de serveurs mainframes, associés à un OS spécifique (System z, GCOS) ou à une application spécifique (bases de données par ex. SAP HANA). Ce type d'architecture privilégie la consistance des transactions à la performance de calcul.

La seconde préside aux destinées du calcul intensif (supercalcul / HPC). Il a d'abord donné lieu à la conception de machines ultra-denses basées CPU et a ensuite vu l'arrivée d'accélérateurs plus spécifiques (manycore, composants reprogrammables FPGA, processeurs graphiques GPU).

DE L'IMPORTANCE CRUCIALE DU RÉSEAU DANS LES ARCHITECTURES

Cette architecture a aussi nécessité le développement de nouveaux réseaux basse latence capables de gérer les congestions.

Aujourd'hui le réseau est partout. Il permet de centraliser le stockage et de garantir l'accès aux données en tout point de l'infrastructure. Il permet d'assurer la communication entre les GPUs et de décharger les processeurs centraux des multiples chargements mémoire. Il s'assure aussi que les données sont transportées de manière sécurisée, ségréguée entre les usages et les utilisateurs.

Il constitue la vraie colonne vertébrale des architectures modernes. Sans un réseau adapté, il n'est pas possible de servir efficacement les cas d'usages.

En 2025, le cloud computing cherche à fédérer tous les usages. En parallèle le supercalcul, qui était jusqu'à

présent de la « haute couture », cherche à s'industrialiser. Comme en géopolitique, les frontières se troublent, et les architectures hybrides sont désormais indispensables.

L'un comme l'autre était jusqu'à présent incapables d'orchestrer des ressources et en même temps de distribuer efficacement des tâches de calcul.

La conteneurisation vient changer cela et constitue l'abstraction capable de fédérer les cas d'usage. Le container est transportable. On peut l'orchestrer. On peut planifier l'exécution de tâche. Et il peut également disposer d'une grande mémoire.

Les architectures hybrides modernes sont celles qui associent l'abstraction container à des îlots de calcul de différente nature :

- General purpose ;
- Calcul intensif ;
- GPUs ;
- FGPA, QPUs, machines SMP...

Le tout en utilisant le réseau pour assurer un échange de qualité entre stockage et calcul, permettant l'accès direct aux composants (GPU direct, voire composants désagrégés) et en réalisant au passage une partie du traitement de la donnée.

Certains îlots sont chargés d'assurer la persistance de la donnée, et donc sa protection, d'autres doivent pouvoir se comporter comme des dispositifs de calcul transitoire sans persistance, mais avec une capacité de reprise. Ces principes ouvrent la voie au cloud étendu jusqu'à l'Edge.

L'IA EN EMBUSCADE AGIT COMME CATALYSEUR DU CHANGEMENT

L'avènement de l'IA, cas d'usage mixte entre supercalcul (entraînement des modèles), cloud computing (inférence des modèles) et SMP (bases de données) accélère l'adoption de ces architectures hybrides.

Le Sauron des data centers n'est pas si loin, même s'il ne s'agit pas d'un anneau unique mais de la collaboration de tous les anneaux entre eux pour constituer le futur de nos architectures informatiques. ■



NVIDIA : OSER ÊTRE UN PIONNIER

Jensen Huang
CEO, NVidia

Nous avons rencontré Jensen Huang, CEO de NVidia pour partager sa vision des nouvelles technologies numériques, de leurs usages, de leurs évolutions futures et de leur impact sur la société. Ainsi que sur la façon dont NVidia va devoir se réinventer pour y faire face.

L'IA, UNE DÉCENNIE FONDATRICE ET UNE RÉVOLUTION À VENIR

Au cours des dix dernières années, nous avons assisté à la naissance de l'IA moderne. Elle est passée de l'IA perceptuelle à l'IA générative, puis à l'IA raisonnante, et nous sommes aujourd'hui à l'aube de l'IA physique. À chaque étape, les capacités des ordinateurs se sont élargies, bouleversant la santé, l'industrie ou encore la recherche, et ouvrant la voie à des usages totalement inédits.

Si l'on se projette dans les dix prochaines années, nous sommes à l'aube d'une nouvelle révolution industrielle. L'IA sera partout : dans chaque entreprise ; chaque société ; et jusque dans notre quotidien. Tout comme l'électricité est devenue l'infrastructure de la révolution industrielle précédente, les « usines d'IA » seront celle de la prochaine. Elles alimenteront les industries mondiales en « intelligence », qui sera pour elles une nouvelle ressource essentielle.

Et le plus passionnant, c'est qu'à l'avenir, chacun de nous pourra être accompagné, augmenté et amélioré par l'IA. Pour la première fois, une nouvelle technologie offre la possibilité d'élever, de manière égalitaire et simultanée, tous les individus, partout dans le monde. Si, dans le passé, la technologie était difficile à utiliser, l'IA change la donne. Tout le monde peut en bénéficier.

HPC ET IA : MOTEURS D'UNE TRANSFORMATION UNIVERSELLE

Le calcul haute performance (HPC) et l'Intelligence Artificielle sont des technologies révolutionnaires. Elles ne se contentent pas d'accélérer l'existant, elles créent des industries applicatives entièrement nouvelles. IA agentique, IA raisonnante, IA robotique...

Autant de technologies appelées à transformer l'ensemble des secteurs, de la santé à l'éducation, de la mobilité à l'industrie manufacturière. Chaque industrie sera révolutionnée.

Aujourd'hui, certains s'inquiètent : si nous automatisons trop, la productivité ne se traduira-t-elle pas simplement par une diminution du nombre d'emplois ? L'histoire nous montre le contraire. Au cours des soixante années de l'ère informatique, lorsque la productivité a augmenté, l'emploi a également augmenté. Pourquoi ? Parce que lorsque nous avons plus d'idées pour un avenir meilleur, c'est la productivité qui nous permet de les concrétiser.

L'IA n'est pas l'apanage de quelques personnes ou de quelques pays, elle est pour tout le monde. Elle générera de nouveaux métiers, de nouvelles entreprises, de nouvelles industries. Elle fera progresser la société dans son ensemble. Pour la première fois, notre industrie a créé une technologie si utile et si universelle, qu'elle touche tous les domaines, des arts aux sciences de la vie, du service à la clientèle aux sciences physiques, des médias à la fabrication. Combinés, l'IA et le HPC ne représentent pas seulement une nouvelle vague d'innovation, mais le début d'une nouvelle révolution industrielle à part entière.

L'AVENIR SERA QUANTIQUE... ET HYBRIDE

L'ensemble du secteur reconnaît que le futur du calcul reposera sur une informatique hybride mixant approches classique et quantique. Loin de l'image de l'ordinateur quantique isolé dans un coin de laboratoire, la magie opère lorsque vous le connectez avec les supercalculateurs GPU pour en tirer la quintessence. Ces derniers assurant le contrôle, la correction d'erreurs et l'orchestration des opérations.

Les progrès réalisés en matière de correction d'erreurs ouvrent des perspectives déterminantes. Et je peux vous dire que tous les centres de calcul intensif que j'ai rencontrés se tournent sans exception vers le





NVIDIA: OSER ÊTRE UN PIONNIER

Jensen Huang
CEO, NVidia

• • •

mixte quantique-classique. L'avenir sera donc bien quantique, mais toujours en symbiose avec le calcul classique, soutenu par les supercalculateurs GPU.

VERS L'IA PHYSIQUE ET LA ROBOTIQUE GÉNÉRALISÉE

Tout ce qui bouge deviendra un jour autonome. Nous entrons dans l'ère de l'IA physique, une IA qui comprend les lois de la physique, la gravité, l'inertie, les relations de cause à effet, la permanence des objets, et un certain bon sens appliqué au monde réel.

Cela signifie que toutes les entreprises manufacturières auront besoin de deux usines. Par exemple, un constructeur automobile aura une usine pour produire physiquement ses voitures, et une usine d'IA pour créer l'intelligence qui les animera. Cette idée de deux usines, l'une pour le produit physique, l'autre pour son intelligence numérique, est l'avenir de toutes les industries.

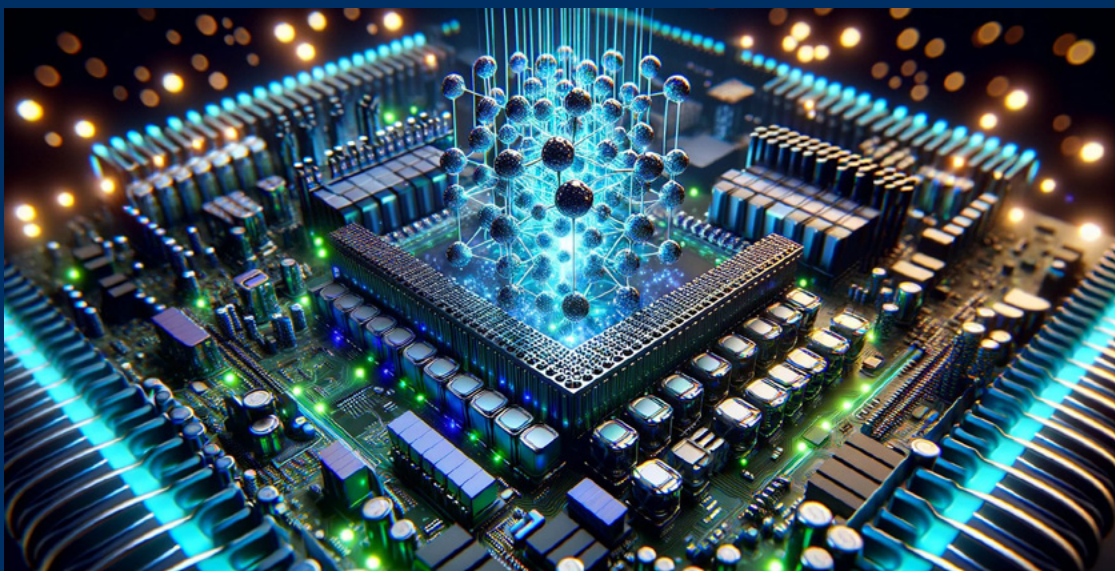
Dans un contexte de pénurie mondiale de main-d'œuvre, il nous manque 30 à 40 millions de travailleurs

qualifiés, la robotique prendra le relais. Elle permettra aux industries de se développer, aux économies de croître et aux sociétés de prospérer. C'est pourquoi je pense que le « moment ChatGPT pour la robotique » est imminent. Les technologies sont en place (simulation, formation, tests, déploiement dans le monde réel) et nous assisterons très bientôt à des avancées révolutionnaires dans le domaine de la robotique générale, qui nous surprendront tous.

NVIDIA ET L'ART D'OSER EN PIONNIER

Depuis toujours, NVidia se distingue par sa volonté d'affronter les défis que personne d'autre n'ose relever. C'est ainsi que nous abordons chaque avancée majeure, non par bravade, mais avec curiosité et conviction.

Qu'il s'agisse de l'IA physique, de la robotique ou du calcul hybride quantique-classique, NVidia restera fidèle à son esprit fondateur: oser être la première, faire les bons choix et apporter une contribution suffisamment significative pour donner un sens à ses équipes et faire progresser le monde. ■



NVIDIA Quantum Cloud © NVidia

LES DÉFIS DU STOCKAGE : ÉVOLUER DANS UN PAYSAGE FRAGMENTÉ ET INNOVER À L'ÈRE DE L'IA

Jean-Thomas Acquaviva

Chef d'équipe Groupe de recherche, DDN Storage

La numérisation du monde se poursuit, poussée par une soif inextinguible d'Intelligence Artificielle et la généralisation des dispositifs d'acquisition de données : de l'IoT aux smartphones, en passant par les voitures connectées et les grands instruments scientifiques. Cette croissance continue des données remodèle le paysage des technologies de stockage, remettant en question les paradigmes et les normes établies. La décennie à venir promet une révolution dans la manière dont nous stockons, accédons et traitons les données. Une transition fondamentale est en cours, depuis une approche centrée sur l'humain et basée sur les fichiers vers un continuum plus fragmenté, axé sur les communications machine-à-machine.

Historiquement, les systèmes de stockage étaient centrés sur l'humain, organisés autour des systèmes de fichiers et de la norme POSIX. Cependant, cette époque est révolue. Le stockage objet a déjà conquis une part significative du marché, et nous anticipons une croissance des flux machine-à-machine avec une prolifération de formats de données spécifiques. Cette transition d'une norme vers un monde fragmenté nécessite une réévaluation des stratégies de stockage. Il est hautement prévisible que, suivant l'adage « le logiciel dévore le monde », le logiciel dévorera le stockage et occupera une place prééminente dans toutes les futures solutions de données.

DYNAMIQUES TECHNIQUES ET DE MARCHÉS EN MUTATION

Les technologies de substrats de stockage évoluent continuellement, mais ne disparaissent pas ; elles s'empilent. Si les disques flash supplantent les disques durs en termes de vitesse et de densité, les disques durs restent une solution viable pour les données froides. De surcroît, il existe un marché en expansion pour l'archivage à zéro-énergie, aujourd'hui autour des bandes magnétiques, demain des disques durs ou même de l'ADN. L'industrialisation des mémoires persistantes, un temps stoppée par l'expérience ratée d'Optane, reprendra dans un avenir proche. Toutes ces technologies resteront masquées aux utilisateurs finaux, car le stockage cloud est devenu la force dominante, ne représentant plus qu'une abstraction aux applications.

TENDANCES ÉMERGENTES ET CONTINUUM DE DONNÉES

Parmi les grandes tendances prévisibles pour les dix prochaines années figure le stockage intégré à la périphérie (edge-storage). Un autre changement à venir est l'atténuation progressive de la frontière entre les systèmes de stockage et les technologies de réseau pour créer un « continuum de données ». Un concept crucial au sein de ce continuum est la gravité des données, selon lequel les données attirent les données et les services attenants. La gestion efficace de la gravité des

LES DÉFIS DU STOCKAGE : ÉVOLUER DANS UN PAYSAGE FRAGMENTÉ ET INNOVER À L'ÈRE DE L'IA

données exige des solutions de stockage innovantes capables de gérer et de transférer les données dans des environnements de plus en plus distribués. La gravité des données pousse également à la désagrégation du stockage, au stockage computationnel, c'est-à-dire l'intégration de capacités de calcul au sein des dispositifs de stockage.

La pleine réalisation du continuum requiert une logistique des données efficace et robuste. Or, cette dernière n'est pas uniquement un problème technique, mais elle implique des transformations au niveau des organisations, ce qui constitue un frein important à son développement.

SÉCURITÉ ET SÛRETÉ À L'ÈRE POST-EXASCALE

À mesure que nous avançons vers un monde de plus en plus interconnecté, la cybersécurité devient primordiale. La protection des données contre les accès non autorisés, les cybermenaces, exige des stratégies de cybersécurité efficaces, utilisant le chiffrement, des contrôles d'accès rigoureux et des mécanismes avancés de détection d'intrusion. Cette « cybersécurité pour un monde ouvert » est une composante essentielle du continuum de données.

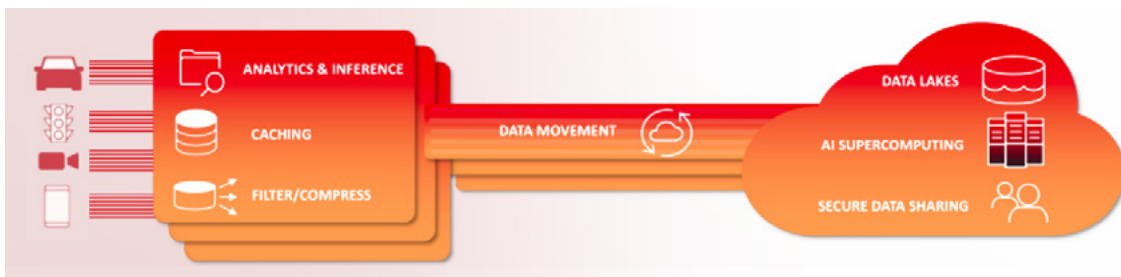
La relation symbiotique entre l'IA et le stockage s'approfondit également : « Le stockage pour l'IA, l'IA pour le stockage ». Les outils de surveillance et d'analyse basés sur l'IA deviendront indispensables pour gérer et optimiser les systèmes de stockage complexes. Ces outils permettront l'analyse des données en temps réel, la maintenance prédictive et la résolution automatisée des problèmes, améliorant considérablement la fiabilité.



POST-EXASCALE ET CONTINUUM DE DONNEES

Notre vision pour l'ère post-exascale des solutions de stockage est celle d'un paysage fragmenté, caractérisé par l'absence de normes universelles, où les mémoires flash joueront un rôle prédominant. Nous anticipons un continuum de données fluide, où la gravité des données dicte la proximité avec le calcul, avec une cybersécurité avancée et intégrée nativement. Ce continuum implique une certaine perte de singularité pour le secteur du calcul haute performance, celui-ci devant s'intégrer dans un écosystème technologique plus large et plus diversifié.

En favorisant la collaboration, en adoptant des benchmarks ouverts et en abattant les barrières entre les start-up de l'IA et la communauté HPC, l'Europe peut stimuler l'innovation. Avec leurs riches écosystèmes scientifiques et industriels, la France et l'Europe disposent de ressources significatives : génération de données, expertise humaine et infrastructures de premier ordre. La capacité de nos institutions à jouer sur le long terme, illustrée par Teratec, qui célèbre 20 ans d'activité, nous place, pour une fois, en position de leader dans cette nouvelle course. ■



AVENIR DU STOCKAGE MASSIF DE DONNÉES POUR HPC ET IA

Gary Grider

Deputy Division Leader MS B260 HPC, Los Alamos National Laboratory

Le domaine du stockage HPC a connu des changements récents, mais pas grand-chose que l'on puisse qualifier de révolutionnaire. Les progrès en performances – bande passante et capacité – du stockage flash exercent une pression sur les marchés du stockage sur bande magnétique. De nouveaux types de supports de stockage à long terme, tels que la céramique et l'ADN, émergent. Bien entendu, le stockage orienté objet suscite également une activité importante et ressemble de plus en plus au stockage à base de fichiers à bien des égards, les humains semblant avoir besoin de hiérarchie.

DE L'IMPORTANCE CROISSANTE DES DONNÉES

Le changement principal est l'évolution du marché vers l'informatique centrée sur les données. Avec l'avènement de l'IA, ce ne sont pas seulement les sites HPC classiques qui ont besoin de technologies de stockage évolutives ; tous ceux qui s'engagent dans l'IA ont besoin d'une solution de stockage pouvant passer à l'échelle. D'où l'apparition d'un plus grand nombre de systèmes de stockage commerciaux évolutifs mais avec des interfaces propriétaires. Les cas d'usage d'entraînement de modèles d'IA ont rendu le stockage local sur les nœuds de calcul populaire. À bien des égards, cette tendance économique centrée sur les données a nuí aux solutions communautaires ouvertes.

La plus grande surprise est l'émergence de Parallel NFS (PNFS) qui avait été lancé en 2002 pour tenter de créer un client de système de fichiers parallèle

commun à tous les systèmes d'exploitation. La communauté et l'industrie ont progressé lentement et le système est aujourd'hui en passe de devenir le meilleur de sa catégorie pour de nombreuses charges de travail évolutives, telles que l'écriture de nombreux fichiers vers de nombreux fichiers ou un seul fichier, les métadonnées évolutives, les tempêtes de lecture, la distance, etc. La communauté PNFS s'attaque même à l'effacement côté client avec une nouvelle approche qui devrait rendre l'écriture d'effacement non alignée côté client aussi efficace que l'effacement côté serveur, ce qui mettrait fin à la nécessité de paires de serveurs de basculement. Quatre entreprises vendent ou travaillent activement sur PNFS. Si la communauté PNFS réussit, un système de stockage évolutif de pointe comprenant des clients, des serveurs de métadonnées et des serveurs de données pourrait être disponible via la communauté ouverte et distribué avec votre système d'exploitation.

VERS DES TRAITEMENTS AU PLUS PROCHE DES DONNÉES

Une autre nouveauté importante consiste à pousser le traitement à proximité du support de stockage. L'indexation, l'intégration de vecteurs et d'autres traitements centrés sur les données sont effectués plus efficacement à proximité du stockage des données. De nombreuses solutions tirent parti du riche écosystème Apache Analytics, qui comprend des formats communs pour les données au repos, en mémoire et en vol, ainsi

que des normes pour les techniques de poussée des requêtes pour les réductions à proximité du stockage.

L'utilisation des sciences de l'information va changer notre monde dans la prochaine décennie. Malgré l'incertitude sur les charges de travail, du nouvel ordre mondial selon lequel les usines d'informatique en nuage et d'IA sont les moteurs du volume dans les secteurs de l'informatique, des réseaux et du stockage, deux directions semblent évidentes. La communauté HPC/AI au sens large a besoin

d'alternatives à l'utilisation de ressources de type cloud et AI factories et d'outils commerciaux propriétaires. Disposer de solutions communautaires reposant sur des logiciels libres/standards et tirer parti de vastes écosystèmes ouverts semble impératif. Étant donné que l'avenir repose sur les données, qu'il s'agisse de fonds existants ou de nouvelles sources, il est également évident qu'il faut des outils riches et sécurisés pour gérer et augmenter la valeur des données. ■

■ ADN, LA MÉMOIRE DU FUTUR

Marc Antonini

Directeur de Recherche, CNRS

La mémoire de l'humanité repose sur notre capacité à gérer des volumes de données en croissance continue. D'ici 2040, plusieurs milliers de zettaoctets (10^{21} octets) devront être conservés sur des périodes allant de quelques années à plusieurs siècles. Une telle masse d'information représente un défi inédit : les technologies actuelles de stockage apparaissent déjà insuffisantes. Leur déploiement massif dans des centres de données nécessiterait des ressources énergétiques considérables et soulèverait de graves problèmes liés à l'empreinte carbone, à la maintenance et à l'obsolescence rapide du matériel.

Dans ce contexte, l'une des alternatives les plus prometteuses repose sur le stockage moléculaire, et en particulier sur l'ADN synthétique. Cette molécule présente des atouts uniques : une densité de stockage extrêmement élevée, une stabilité sur de très longues périodes et des conditions de conservation relativement simples. Elle offre ainsi une voie crédible pour concevoir des solutions durables, sobres en énergie et capables de répondre aux besoins de préservation des données de demain.

Le PEPR exploratoire MolecuArXiv s'inscrit pleinement dans cette perspective. Son objectif est clair : développer, à l'échelle de cinq ans, les recherches et infrastructures nécessaires pour accélérer d'un facteur 100 la vitesse actuelle du cycle d'écriture/lecture sur ADN, tout en divisant par 100 son coût. Atteindre un

tel objectif permettrait de franchir un cap décisif vers l'industrialisation du stockage moléculaire.

Ce programme, inscrit dans le cadre de France 2030, a été confié au CNRS pour la direction scientifique, l'ANR assurant le rôle d'opérateur. Doté d'un budget de 20 m€ sur 7 ans, il mobilise une communauté scientifique pluridisciplinaire réunissant des expertises en chimie, biologie, informatique, microfluidique et nanotechnologies. Quatre projets ciblés ont été lancés fin 2022 pour poser les bases de cette nouvelle technologie. En 2025, trois nouveaux projets sélectionnés par appel à projets sont venus enrichir cette dynamique et élargir encore le spectre des recherches menées.

Si l'ADN constitue l'axe principal de MolecuArXiv en raison de sa densité d'information et de sa stabilité éprouvée, le programme explore également des alternatives, notamment des polymères synthétiques. Ceux-ci pourraient apporter des avantages complémentaires, comme une vitesse d'écriture accrue ou une meilleure évolutivité pour certaines applications spécifiques. Cette approche duale, centrée sur l'ADN tout en ouvrant la voie à d'autres polymères, permet de couvrir un large éventail de verrous scientifiques et technologiques.

En réunissant des équipes issues de disciplines variées, MolecuArXiv vise à accélérer l'émergence d'une technologie de rupture, appelée à transformer durablement nos modes de conservation de l'information. Ce programme ambitionne de construire un leadership français et de positionner la recherche académique nationale au meilleur niveau international, dans un domaine stratégique pour l'avenir de la société numérique. ■

VIRTUALISATION ET ORCHESTRATION

Guillaume Colin de Verdière

Directeur du programme transversal de compétences simulation numérique, CEA

Gilles Wiber

Chef de département adjoint, CEA

Jean-Philippe Nominé

Fellow, CEA

Depuis dix ans la pression des utilisateurs et des technologies du cloud a contribué à une évolution des modalités d'usage des supercalculateurs. Les machines ne sont plus vues comme des monolithes, mais comme des ensembles plus flexibles de ressources. Ainsi des logiciels de virtualisation et d'orchestration sont de plus en plus déployés sur les supercalculateurs.

VIRTUALISATION ET CONTENEURISATION

La virtualisation (et sa variante dite conteneurisation) rend les applications indépendantes des ressources matérielles et/ou de l'environnement système. Les machines virtuelles reproduisent une machine physique complète. Les conteneurs, plus légers, embarquent une application et toutes ses dépendances logicielles ; construits une fois, ils permettent ensuite des déploiements sur des machines diverses. Ces approches ont de nombreux avantages pour les utilisateurs, mais aussi pour les administrateurs système.

Pour un administrateur, ces approches permettent de tester des montées de version du système d'exploitation, ou réserver tout ou partie de la machine pour des expérimentations, sans perturber la production. Ils permettent aussi d'encapsuler les tâches utilisateurs dans un environnement sécurisé, n'exposant que le strict minimum des ressources requises, voire octroyer des droits localement accrus sur les ressources. Dans le cas de matériels particuliers (carte graphique ou réseau propriétaire notamment), il sera nécessaire de faire l'adaptation spécifique avec les pilotes ad hoc.

Pour un utilisateur, se placer au sein d'un conteneur (un espace par exemple Docker ou Kubernetes) permet d'exécuter une application dans un environnement totalement maîtrisé, quelles que soient les versions du système hôte natif de la machine d'accueil. Il aura la garantie de la protection de ses données et la possibilité de déplacer son code d'une machine à l'autre, selon ses besoins ou les ressources qui lui seront allouées, tout en ayant une garantie de performance matérielle.

Ces méthodes sont déployées sur le supercalculateur EXA1 du CEA/DAM, grâce au logiciel open source PCOCC* (Private Cloud On a Compute Cluster). PCOCC sait gérer et instancier soit des conteneurs soit des machines virtuelles avec les mêmes interfaces vues de l'utilisateur. L'introduction de conteneur permet aussi d'isoler des travaux de calcul les uns des autres et ainsi accroître la sécurité du centre de calcul (on arrive à isoler jusqu'au niveau du réseau). Ces travaux vont se poursuivre pour mise en œuvre dans le futur calculateur exaflopique EuroHPC français (Alice Recoque) qui sera installé au CEA/TGCC.

ORCHESTRATION ET WORKFLOWS

Par ailleurs, de plus en plus d'études à base de simulation et analyse de données massives requièrent des phases de traitement diversifiées, enchaînant souvent différents codes de calcul avec des étapes de pré et post-traitements, éventuellement alternées. Allant bien au-delà des allocations et ordonnancements rigides du HPC traditionnel, **l'orchestration** de

* <https://github.com/cea-hpc/pcocc>

VIRTUALISATION ET ORCHESTRATION

ces différents traitements conduit à des *workflows* complexes ; de nombreux outils et API sont développés à cet effet, sans que se dégage à ce jour de grand standard en la matière. Les gestionnaires de *workflows* peuvent présenter des interfaces graphiques *low-code*, *no-code* qui vont décrire l'ensemble des traitements souhaités par exemple pour un jumeau numérique.

Virtualisation et orchestration contribuent à une démocratisation de l'usage du HPC car il ne sera plus nécessaire d'être un spécialiste des supercalculateurs

et de leur environnement pour être productif, dès lors qu'une abstraction convenable du centre de calcul et de ses ressources est définie et outillée.

Les communautés des chercheurs en IA, des utilisateurs de réseaux de neurones, qui ne viennent pas forcément du monde du HPC traditionnel, mais également les industriels, sont particulièrement en attente de ces méthodes. Pour répondre à leur besoin, le recours aux logiciels de virtualisation et d'orchestration devient une nécessité pour leur apporter flexibilité, efficacité et sécurité. ■



Sur cette allégorie, les murs d'une cuisine représentent les circuits d'un ordinateur. L'étagère contient des mini-cuisines (des machines virtuelles : leurs ustensiles représentent les outils de la virtualisation, au-dessus des ressources de bas niveau) et des boîtes de repas (les repas tous faits représentent les applications qui sont dans les conteneurs). Au centre, un opérateur d'orchestration coordonne l'enchaînement d'activités - différents modules - pour que des tâches globales en résultent de façon cohérente. (Image générée par l'IA Gemini).

HPC ET EXASCALE : LA PUISSANCE À L'ÉPREUVE DE LA COMPLEXITÉ LOGICIELLE

Luc Giraud

Directeur de recherche, Centre Inria de l'université de Bordeaux, Airbus CR & T, CERFACS

Brice Goglin

Directeur de recherche, Centre Inria de l'université de Bordeaux, LaBRI

L'arrivée des premières machines exascale en 2022 est l'aboutissement d'une longue course technologique engagée au début des années 2000. Dès le départ, les concepteurs de ces machines ont compris que le principal obstacle serait l'énergie. Cela a dicté des choix radicaux : les processeurs graphiques (GPU) ont remplacé les processeurs classiques (CPU) pour leur rendement supérieur ; le nombre d'unités de calcul a explosé ; la réduction des mouvements de données est devenue une obsession, car transporter une donnée entre deux nœuds peut désormais coûter plus d'énergie que de la recalculer sur place.

DES PLATEFORMES TOUJOURS PLUS COMPLEXES ET DES ENJEUX STRATÉGIQUES

La transition vers l'exascale s'est accompagnée d'une profonde transformation des architectures matérielles. Les calculateurs modernes reposent sur des combinaisons hybrides de CPU et de GPU, parfois des mémoires hétérogènes, etc.

Les piles logicielles reposent souvent sur l'environnement CUDA du fournisseur NVIDIA. Les alternatives n'étant pas encore assez matures, le risque de dépendance technologique vis-à-vis de quelques acteurs est réel, fragilisant l'autonomie stratégique des industriels et des États.

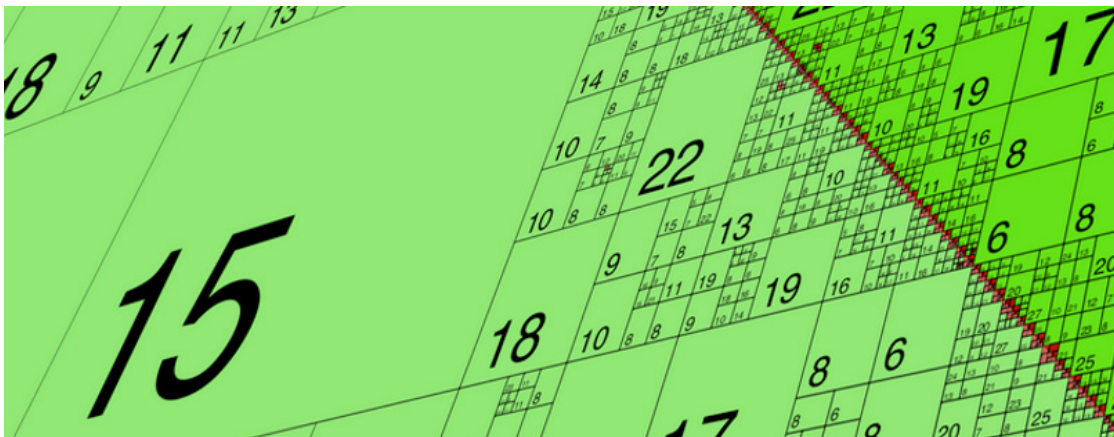
L'écosystème peine à suivre : les briques logicielles sont multiples, hétérogènes, souvent propres à certains matériels. Cela fragilise la portabilité des applications.

Ce risque est partiellement compensé par l'allongement de la durée de vie des machines pour des raisons environnementales et financières, mais impose en retour de garantir la stabilité de l'ensemble des briques logicielles — compilateurs, bibliothèques, environnements d'exécution, applications — sur des périodes prolongées, malgré leurs évolutions asynchrones.

Les standards logiciels se font encore attendre. Le vénérable protocole MPI demeure encore aujourd'hui le socle commun incontournable pour la communication entre nœuds. En revanche, pour la gestion fine des tâches ou des graphes d'exécution, aucun consensus n'a encore émergé. Des initiatives comme OpenMP progressent, tandis que les modèles fondés sur la programmation par graphes de tâches suscitent un intérêt croissant. Encore complexes à mettre en œuvre, elles laissent entrevoir la possibilité d'une standardisation future qui pourrait faciliter l'indépendance vis-à-vis des solutions propriétaires des constructeurs.

DES APPLICATIONS SCIENTIFIQUES EN PLEINE MUTATION

Les applications qui exploitent ces supercalculateurs ne cessent de se complexifier. Les codes numériques d'aujourd'hui intègrent des phénomènes multiphysiques, multi-échelles et parfois temporels, rendant la simulation quatre dimensions une nécessité. Dans des domaines comme la mécanique quantique ou l'astrophysique, les modèles explorent même des espaces d'états bien plus vastes.



Zoom sur une matrice hiérarchique - University of Bayreuth, Germany

Autre transformation majeure : l'irruption de l'incertitude. Les grandes simulations intègrent de plus en plus des approches de quantification d'incertitudes ou d'assimilation de données, pour croiser les calculs numériques avec les données d'observation expérimentales. Un unique calcul exascale déterministe a généralement peu d'intérêt : c'est la réalisation massive d'ensembles de calculs variés qui devient la norme.

Cette sophistication se heurte à une autre difficulté souvent intrinsèque : l'irrégularité des données et par conséquent des algorithmes qui les manipulent. La charge de calcul n'est plus uniformément répartie entre les processeurs ; certains nœuds patientent pendant que d'autres s'embourbent dans des tâches complexes. L'équilibrage dynamique de la charge, le vol de travail d'un nœud à l'autre ou encore l'ordonnancement intelligent des graphes de tâches issus d'une expression haut niveau d'implantation moderne d'algorithmes deviennent des éléments indispensables du calcul efficace.

INTELLIGENCE ARTIFICIELLE : UNE RÉVOLUTION ANNONCÉE DANS LE HPC

L'irruption de l'intelligence artificielle (IA) dans le monde du calcul scientifique est sans doute la transformation la plus inattendue de ces dernières années. Les accélérateurs matériels conçus pour l'IA imposent

des formats numériques nouveaux (FP16, BFLOAT16), bien plus efficaces énergétiquement que la double précision classique (FP64) chère aux ingénieurs et physiciens. À terme, il est possible que ces formats deviennent dominants, sauf dans les calculs où la précision extrême reste indispensable.

Au-delà du matériel, l'IA promet aussi d'accélérer le calcul lui-même : en construisant des modèles réduits capables d'approcher des simulations coûteuses, en optimisant les solveurs numériques ou en prédisant dynamiquement la meilleure répartition des tâches. Mais ces usages restent pour l'instant préliminaires mais prometteurs dans le domaine des grandes simulations industrielles, où l'explicabilité mathématique peut encore primer sur les approches de l'IA.

AU-DELÀ DE LA PUISSANCE : LES DÉFIS SILENCIEUX DU CALCUL EXASCALE

L'exascale n'est pas une simple montée en puissance des ordinateurs existants : il s'agit d'une rupture complète dans la manière de concevoir, de programmer et d'exploiter le calcul scientifique. Cette révolution silencieuse impose aux chercheurs, aux ingénieurs mais aussi aux décideurs d'anticiper des choix technologiques qui engageront les capacités industrielles et scientifiques des décennies à venir. ■

HPC ET INTELLIGENCE ARTIFICIELLE : INNOVATIONS ET CONVERGENCE

Gabriel Antoniu

Directeur de recherche, responsable scientifique de l'équipe KerData, Inria

Bruno Raffin

Directeur de recherche, responsable scientifique de l'équipe DataMove, Inria

Olivier Beaumont

Directeur de recherche, responsable scientifique de l'équipe Topal, Inria

L'intelligence artificielle (IA), et en particulier l'apprentissage profond, a transformé de manière disruptive le domaine du calcul haute performance (HPC). Ces deux disciplines, autrefois distinctes, convergent aujourd'hui pour répondre à des défis scientifiques et industriels d'une complexité croissante. Cette convergence est mise en évidence dans des agendas stratégiques pour la recherche européenne et fait actuellement l'objet de riches échanges au sein de groupes de travail internationaux tels qu'InPEX. Le HPC, traditionnellement fondé sur la résolution numérique d'équations complexes comme les équations aux dérivées partielles, tire parti des avancées de l'IA pour accélérer les calculs, optimiser les modèles et explorer de nouveaux paradigmes computationnels. À l'inverse, l'IA s'appuie sur les infrastructures HPC pour entraîner des modèles de plus en plus volumineux, nécessitant une puissance de calcul et une gestion des données à grande échelle. Cet article évoque des développements récents à l'intersection de ces domaines, en mettant en lumière les approches méthodologiques, les innovations algorithmiques et les évolutions des infrastructures qui redéfinissent le paysage du calcul scientifique et ses applications.

APPRENTISSAGE PROFOND ET SIMULATION NUMÉRIQUE

Les réseaux de neurones (RN), au cœur de l'IA, transforment la simulation numérique sur laquelle repose le calcul haute performance (HPC). Contrairement aux

méthodes traditionnelles, qui approchent la solution d'équations complexes, comme les équations aux dérivées partielles (EDP) par discrétisation et calculs itératifs, les RN apprennent à partir de données pour construire des fonctions.

Des RN sont combinés avec des solveurs classiques d'EDP. Le solveur appelle le RN, qui implémente un noyau de calcul construit par apprentissage à partir de données. Cet appel peut être exécuté localement ou déporté sur GPU. Le gain de précision obtenu permet au solveur de travailler à une résolution plus faible, réduisant le temps de calcul.

Des RN sont entraînés pour se substituer complètement aux solveurs en apprenant à partir des données qu'ils produisent. Le modèle obtenu peut être vu comme un compresseur de données (avec perte) très efficace, doublé de capacités d'interpolation. Compact et rapide, il peut servir à explorer interactivement les données de simulation, ou à être couplé à des capteurs de données comme jumeau numérique. Il peut aussi être utilisé pour l'exploration de l'espace des paramètres à un coût réduit, inaccessible avec le solveur d'origine, qui lui est réutilisé pour valider les solutions prometteuses identifiées par le RN.

Les architectures utilisées, comme les réseaux convolutionnels ou les transformers, proviennent souvent de l'IA classique, mais des modèles spécifiques, comme les PINN (Physics-Informed Neural Networks) ou les FNO (Fourier Neural Operators), sont aussi développés.

Enfin, des techniques avancées, comme la différentiation automatique intégrée aux frameworks de réseaux neuronaux (RN), ont permis le développement d'architectures hybrides combinant étroitement RN et solveurs. Dans ces architectures, le solveur est incorporé au RN et participe au calcul de la sortie. Lors de l'apprentissage, la fonction de perte, concept central des RN, doit être dérivée pour la rétropropagation, une étape clé de l'apprentissage. Cela implique également de dériver le solveur. Grâce aux frameworks de RN, ces dérivées sont calculées automatiquement via la différentiation automatique.

OPTIMISATION DES INFRASTRUCTURES HPC POUR L'APPRENTISSAGE

Les infrastructures de calcul haute performance sont profondément modifiées par l'entraînement de modèles d'intelligence artificielle à grande échelle. La mémoire, le réseau, le stockage et la précision des calculs sont soumis à des contraintes spécifiques. Les architectures ont connu une évolution notable : les processeurs graphiques (GPU) sont devenus des éléments centraux, et les calculs en précision réduite (FP16, bfloat16) se sont généralisés, permettant d'établir un meilleur équilibre entre performance et précision. Parallèlement, les algorithmes d'apprentissage ont intégré des techniques comme la rematérialisation ou la compression, afin de réduire la consommation mémoire.

IMPACT SUR LES RÉSEAUX ET LE STOCKAGE

Les communications collectives jouent un rôle crucial dans l'entraînement distribué qui s'appuie souvent sur plusieurs formes de parallélisme (sur les données, les modèles ou les calculs). Cela entraîne une forte pression sur les réseaux d'interconnexion et nécessite des réseaux spécialisés et des mécanismes permettant de recouvrir efficacement calcul et communication. Du côté du stockage, les grands modèles accèdent généralement séquentiellement à de gros volumes de données, mais les phases de préparation restent exigeantes.

ACCEPTABILITÉ DES APPROXIMATIONS

Contrairement à la simulation numérique, qui produit les données dynamiquement et exige une précision numérique élevée, une forte reproductibilité et des

communications régulières entre processus, l'apprentissage automatique peut tolérer des approximations, des pertes de précision numérique, des communications collectives tronquées ou compressées, ce qui autorise des optimisations plus agressives et une plus grande variété de solutions algorithmiques et de compromis calculs-mémoire.

À terme, on peut sans doute s'attendre à des architectures plus hybrides mêlant calcul et traitement des données en périphérie. Des approches innovantes, comme le calcul en mémoire ou les puces neuromorphiques, pourraient également voir le jour, aux côtés de systèmes spécialisés comme ceux de Cerebras ou Graphcore. En tout cas, l'apprentissage est déjà un moteur central d'évolution pour l'ensemble du calcul intensif.

IA ET HPC : UNE CONVERGENCE DÉCISIVE

La convergence du calcul haute performance et de l'intelligence artificielle marque une étape décisive dans l'évolution des sciences computationnelles. Les progrès de l'apprentissage profond, combinés à la puissance des infrastructures HPC, ouvrent des perspectives inédites pour la simulation numérique, l'optimisation des calculs et l'exploration d'espaces de paramètres complexes. Les réseaux de neurones, qu'ils servent à analyser des données, à augmenter des solveurs traditionnels ou à les remplacer par des modèles de substitution, redéfinissent les approches méthodologiques et logicielles. Parallèlement, l'adaptation des architectures HPC aux exigences de l'IA, avec l'essor des processeurs graphiques, des calculs en précision réduite et des techniques comme la rematérialisation, repousse les limites de l'efficacité computationnelle. À l'avenir, l'intégration croissante de paradigmes hybrides et l'émergence de technologies innovantes, telles que le calcul en mémoire, devraient renforcer cette synergie, permettant de relever des défis scientifiques et industriels toujours plus ambitieux. Cette dynamique consacre l'IA comme un moteur central d'innovation pour le HPC. ■

AGENCE DE PROGRAMME DANS LE NUMÉRIQUE, ALGORITHMES, LOGICIELS ET USAGES

Sophie Proust

Directrice exécutive de l'Agence de programmes dans le numérique, Inria

Créée début 2024 sur décision du Président de la République et confiée à Inria, l'Agence de Programmes dans le Numérique - Algorithmes, logiciels et usages a pour vocation de développer et de piloter des programmes de recherche et d'innovation répondant aux grands défis scientifiques et technologiques dans le domaine du numérique, en s'assurant notamment d'une bonne articulation entre les différents acteurs de la recherche et du tissu économique.

Ses principales missions incluent la construction d'une vision prospective concertée, l'émergence de nouveaux programmes stratégiques, l'établissement de partenariats stratégiques à un fort impact, l'accompagnement de l'écosystème, le transfert des découvertes scientifiques vers des solutions industrielles, la formation et le perfectionnement des compétences des chercheurs et des ingénieurs tout au long de leur carrière, et évidemment la coordination avec les actions européennes et internationales sur le numérique.

L'Agence de programme dans le numérique joue ainsi un rôle clé dans la structuration et la coordination des avancées technologiques en France. En tant que pilier national de l'innovation numérique, elle se distingue par sa capacité à regrouper des programmes liés à des domaines stratégiques, en soutien des stratégies nationales d'accélération, tels que le calcul haute performance (HPC), la cybersécurité, le Système réseau et cloud, l'intelligence artificielle (IA), le quantique, les mondes virtuels ainsi que le numérique en santé, le numérique et l'environnement et le numérique et apprentissages.

UN MODÈLE DÉCLOISONNÉ ET TRANSVERSAL

L'un des principaux atouts de cette Agence réside précisément dans sa capacité à décloisonner les disciplines, en favorisant des interactions riches entre des secteurs qui, historiquement, évoluaient souvent en silos. En regroupant des programmes très complémentaires tels que le calcul haute performance, l'IA, le quantique ou encore le cloud, elle offre une plateforme où les synergies se matérialisent autour de projets novateurs.

Ce décloisonnement permet d'élaborer des objets technologiques transverses qui combinent les forces des différents domaines. Par exemple, le projet AI Factory France porté par GENCI a pu être accompagné par l'Agence au moyen d'une étroite collaboration entre experts de l'IA, du calcul intensif et du cloud.

MIEUX ANTICIPER LES ÉVOLUTIONS SCIENTIFIQUES MAJEURES

À l'heure où l'Europe se dote de supercalculateurs exaflopiques, la communauté doit déjà se préparer à l'ère post-exascale où il faudra composer avec des architectures de calcul spécialisées moins énergivores, et où il faudra répondre à des usages applicatifs mêlant captation et traitement des données, simulation numérique et intelligence artificielle. Ces nouveaux usages appellent à une véritable convergence des méthodes et outils utilisés dans des domaines autrefois cloisonnés. C'est tout l'enjeu du calcul haute performance de demain ! ■



L'IA CHANGE LA FAÇON DE CONCEVOIR

Herbert Taucher

VP for Industrial Research & Pre-development for Integrated Circuits
and Electronics au sein des Foundational Technologies chez Siemens

Nous avons demandé à un grand industriel, qui est aussi éditeur de logiciels de conception, de fabrication et d'exploitation, comment les nouvelles technologies numériques impactent ses activités industrielles et comment il les intègre dans ses logiciels pour les automatiser plus.

Siemens est un grand groupe industriel international, plus de 300 000 personnes dont 50 000 en R&D, avec plus de 10 000 produits au catalogue. Beaucoup sont du matériel toujours plus intelligent, grâce aux logiciels les pilotant. Mais nous avons aussi une large panoplie de logiciels pour la conception (mécanique, électronique, logiciel), la fabrication, ainsi que l'exploitation d'activités industrielles.

Longtemps ingénieur de conception, je suis maintenant responsable de l'activité électronique du groupe, de la recherche fondamentale aux logiciels de conception (EDA). Cela en liaison avec les autres activités du groupe pour produire des systèmes multi-technologies. Je conseille également le CA du groupe pour l'électronique et je suis l'interface sur ces sujets pour la Commission Européenne, les gouvernements locaux et les associations industrielles.

L'IA et encore plus l'IA Générative, vont changer la manière de concevoir les produits Siemens, ainsi que les logiciels de conception associés dans tous les domaines. D'autant que l'on utilise de plus en plus les jumeaux numériques, de la conception à l'exploitation, et que l'on doit faire face à de nouveaux paramètres comme les cybermenaces ou le développement durable. Nos clients nous demandent aussi des produits capables d'évoluer durant leur cycle de vie. Nous devons donc intégrer des spécifications, de la schématique fonctionnelle, de la vérification, du code, de la 3D, des simulations numériques, etc., dans des jumeaux numériques de conception. Puis, les proposer à nos clients sous forme de modèles réduits exécutables à la volée pour simplifier et optimiser l'exploitation de nos produits finis.

LE MATÉRIEL DÉFINI PAR LE LOGICIEL

Pour répondre aux attentes de durabilité de nos clients, nous développons de plus en plus de « matériels définis par le logiciel », c'est-à-dire conçus pour que les fonctionnalités principales reposent sur le contrôle du logiciel, afin de les rendre évolutifs avec le temps pour offrir plus de services aux utilisateurs.

Cela passe par une meilleure compréhension de l'évolution des cas d'usage pour définir le type de degrés de liberté pouvant intéresser nos clients dans le futur. Il s'agit de suppositions éclairées, mais plus vos sources de données et vos modèles sont riches, mieux vous pouvez anticiper de nouvelles fonctionnalités potentielles.

L'IA ACCÉLÈRE LA CONCEPTION

Devant l'ampleur des problèmes à traiter, les concepteurs doivent être soutenus par des données et un ensemble modulaire de modèles intégrés, les aidant non plus à trouver des optimums locaux, mais un optimum global.

L'IA et le HPC devraient nous y aider. On voit déjà apparaître l'IA dans nombre de logiciels de conception, mais avec des niveaux de maturité très différents. Cela va de l'idée novatrice aux démonstrateurs technologiques, qui bien souvent n'ont pas encore été éprouvés à grande échelle. Reste à savoir, s'ils seront adaptés à la complexité et à la taille des problèmes rencontrés, en termes de capacité et de performance.

Les mécanismes d'apprentissage permettent déjà de caractériser de manière multidimensionnelle les paramètres d'un produit, pour en créer un modèle réduit de simulation, tout aussi efficace mais moins gourmand en calcul et plus rapide à exécuter. Nous avons ainsi réduit de trois ordres de grandeur le nombre de simulations Spice nécessaires pour optimiser un produit électronique. Et il en va de même pour la caractérisation des bibliothèques standard ou la simulation analogique. Ce sont les premières applications à bénéficier de l'apport de technologies d'apprentissage automatique et de l'IA.



L'IA CHANGE LA FAÇON DE CONCEVOIR

Herbert Taucher

VP for Industrial Research & Pre-development for Integrated Circuits
and Electronics au sein des Foundational Technologies chez Siemens

Nous développons des outils de conception utilisant l'IA Générative, les graphes de connaissances, les réseaux neuronaux graphiques (GNN), les grands modèles de langage (LLM), etc., pour l'ensemble du portefeuille PLM de Siemens. C'est une remise au goût du jour technologique des systèmes experts, rendue possible par l'explosion des puissances de calcul autorisant des modèles non plus de quelques millions mais de quelques milliards de paramètres.

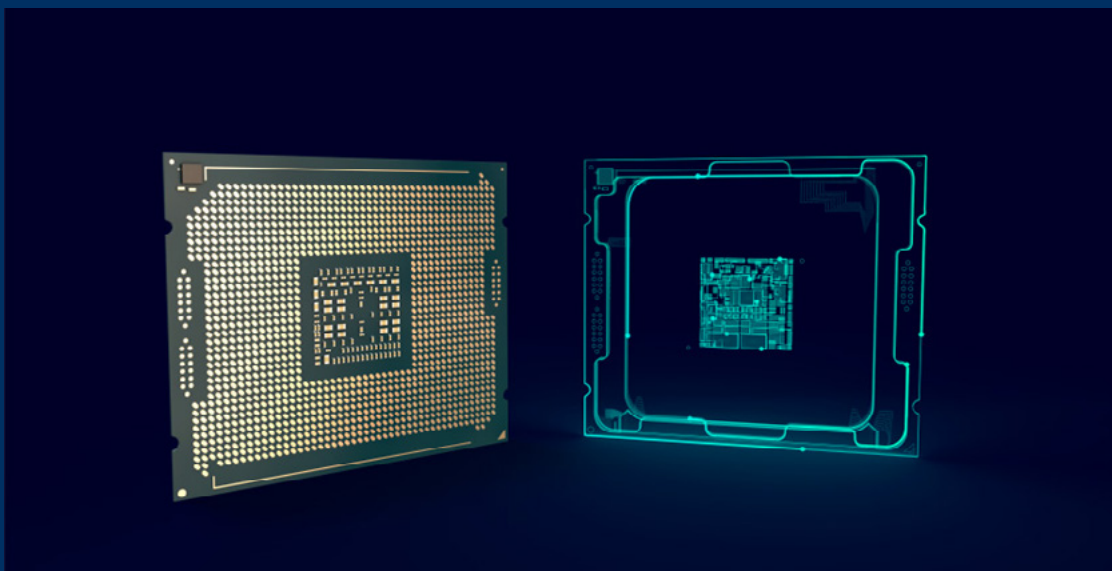
Cette automatisation va aussi nous aider à faire face aux départs en retraite des personnels de conception, de fabrication et d'exploitation, ainsi qu'à la pénurie d'experts qui frappe nos industries. C'est une priorité pour Siemens en tant qu'entreprise, mais aussi pour nos clients.

C'est pourquoi nous avons introduit des copilotes dans nos logiciels, non seulement pour la conception, mais aussi la fabrication et l'exploitation des produits. Cela augmente l'expertise et la productivité des utilisateurs.

LE QUANTIQUE DOIT TROUVER SA PLACE

L'informatique quantique semble intéressante pour certaines utilisations (optimisation logistique, etc.) en stimulant la productivité et la qualité des résultats, mais elle est en concurrence avec d'autres technologies (neuromorphiques...) globalement moins chères. Cela reste un sujet de recherche qui progresse vite, notamment autour du mix classique/quantique. Il reste encore beaucoup de chemin à parcourir tant sur le matériel et la réécriture des logiciels, que sur la compréhension du bon usage, mais nous allons connaître une évolution similaire à celle de l'IA Générative et à l'arrivée des LLM.

À terme, toutes ces technologies devraient nous permettre d'automatiser grandement les processus de conception, dans tous les domaines, en partant d'une vingtaine de lignes de spécifications et quelques valeurs clés. Le concepteur se focalisant plus sur les scénarios d'usage de son produit, ainsi que leur évolution au fil du temps, mais en gardant le pouvoir de décision. ■



Digital Twin Semiconductor image source Siemens

LE CALCUL AU CŒUR DES PRÉOCCUPATIONS DE L'AGENCE DE PROGRAMME ASIC

Jean-Philippe Bourgoin

Directeur de l'Agence de programme ASIC des composants aux systèmes et infrastructures numériques, CEA

Annoncées par le Président de la République, en décembre 2023, les agences de programmes, dont l'agence ASIC confiée au pilotage du CEA, ont été mises en place début 2024 avec pour mission, sur le domaine de l'agence, de construire une vision stratégique, de proposer et mettre en œuvre des programmes nationaux de recherche, plus généralement, de suivre l'ensemble des programmes existants et de mobiliser la communauté académique nationale tout en faisant le lien avec l'industrie.

Le calcul que ce soit au niveau des composants, des systèmes ou des infrastructures numériques est un sujet tout particulièrement considéré par l'Agence ASIC. Il joue en effet un rôle déterminant dans nos sociétés de l'information et de la communication et sa maîtrise est plus que jamais stratégique. En effet les tensions géopolitiques mais aussi les profondes transformations scientifiques et technologiques en cours, notamment le déploiement de l'intelligence artificielle (IA) et la course au quantique, remettent profondément en question les alliances, les dynamiques de développement et les chaînes de valeurs établies.

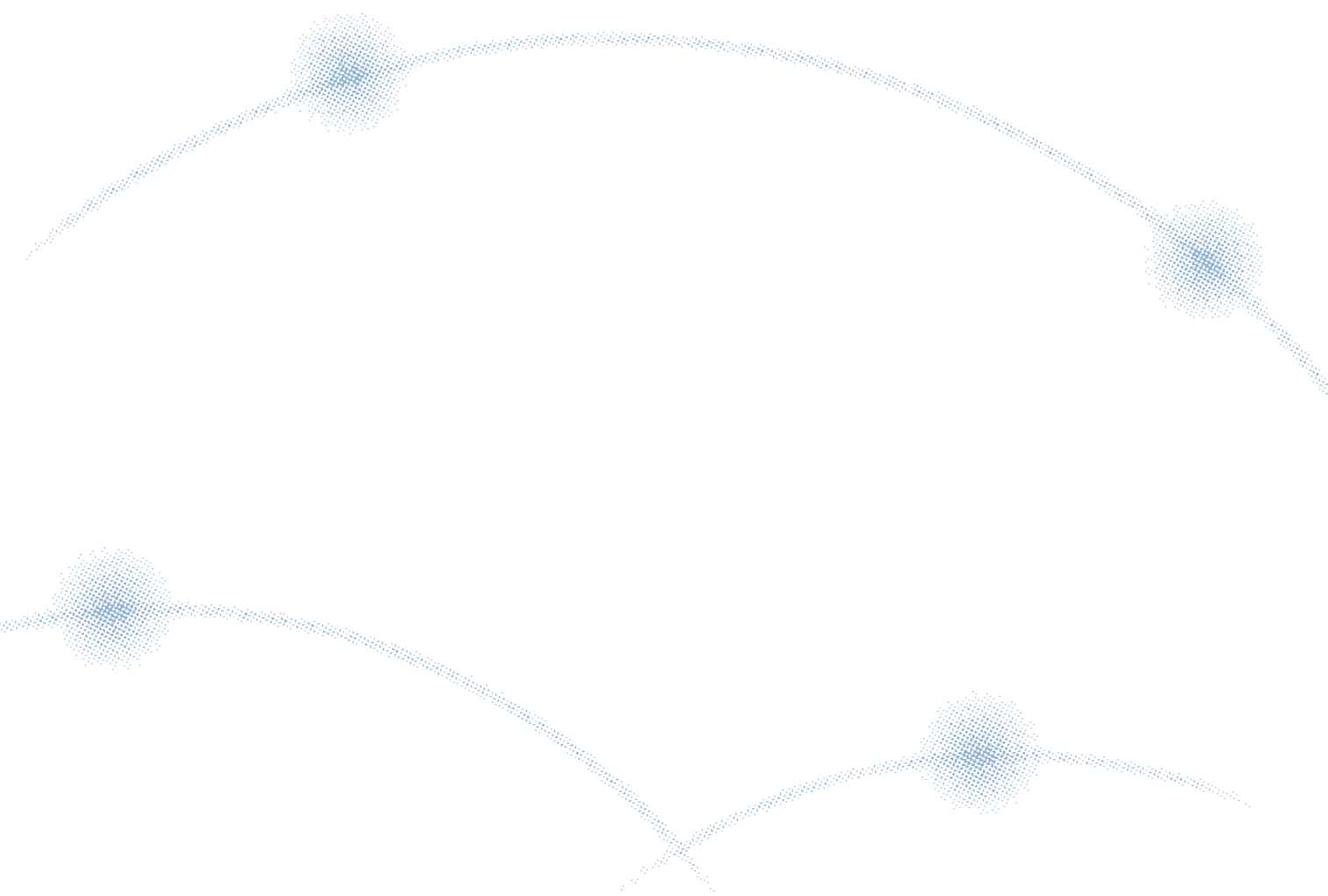
A titre d'exemple, si l'Europe et la France restent très bien positionnées sur l'IA embarquée, elles sont totalement dépendantes pour les composants IA de haute puissance : c'est la raison pour laquelle l'Agence ASIC a proposé à l'État deux programmes. Le premier sur la

conception des composants - en lien avec la plateforme de design européenne se mettant en place dans le cadre du chips act - qui vise à développer la conception de composants à base d'intégration microélectronique 3D, et les outils utilisant les IA génératives. Le second en commun avec l'Agence pilotée par Inria vise par un co-design matériel-logiciel à développer des composants IA ultra-efficacients tirant parti de ruptures comme le calcul dans la mémoire, les approches neuromorphiques, les interconnexions photoniques et l'intégration 3D, pour se replacer dans la course à un horizon de 8 ans.

L'Agence met également en place une réflexion sur les évolutions des infrastructures de calcul - IA et quantique de l'edge au cloud, et les actions à mettre en place pour que la France et l'Europe gardent une maîtrise souveraine de celles-ci.

Il ressort d'ores et déjà des analyses menées par l'Agence qu'au-delà des enjeux scientifiques, technologiques et industriels à maîtriser, l'élargissement du vivier national de compétences, académiques et industrielles est une composante indispensable.

Dans ce contexte, Teratec a depuis 20 ans, montré un exemple remarquable à suivre en matière de sensibilisation, d'acculturation, de structuration et de facilitation de l'accès au calcul sous toutes ses formes. ■







USAGES



INTRODUCTION : USAGES, DE L'INDUSTRIE À L'HOMME

Jean-François Prevéraud
Journaliste

Initialement les technologies de simulation numérique furent réservées à des usages industriels (spatial, aéronautique, automobile, énergie, etc.), pour du calcul de structures ou d'écoulements fluides. Des secteurs qui ont été à l'origine de grands logiciels de conception et codes de calcul (Ansys, Catia, Fluent, Nastran...). Ce sont les outils de CAO (Conception Assistée par Ordinateur), d'IAO (Ingénierie Assistée par Ordinateur), de Mécanique des Fluides, de Simulation Thermique, etc.

Longtemps ces outils propres aux Sciences de l'Ingénieur ont été cantonnés à la validation de produits en fin de phase de conception, car la faible puissance informatique imposait des temps de calcul prohibitifs. Induisant des itérations longues et coûteuses en cas de modification de conception.

DE LA VALIDATION À LA PRÉDICTION

Peu à peu, les technologies numériques ont investi d'autres domaines comme le nucléaire ou la défense qui ont eux aussi développé leurs propres solveurs. Puis ces technologies numériques sont finalement sorties des bureaux d'études pour s'intéresser à la production et à la gestion : FAO (Fabrication Assistée par Ordinateur) ; GPAO (Gestion de Production Assistée par Ordinateur) ; ERP (Planification des Ressources de l'Entreprise), etc. La recherche, le monde de la finance ont aussi commencé à s'y intéresser.

La puissance de calcul se démocratisant et l'offre logicielle sur étagère s'étoffant, tous les secteurs d'activité ont finalement pu bénéficier d'outils de simulation

numérique idoines. Ce qui a aussi permis de faire basculer les usages de la validation à la prédiction. Aujourd'hui on simule avant de concevoir, de fabriquer un produit ou de lancer un service. Et l'arrivée du HPC et de l'IA Générative renforce cette approche.

JUSQU'AU GRAND PUBLIC

Si les grands secteurs industriels sont toujours de grands utilisateurs, une multitude d'autres secteurs ont aussi adopté les outils de simulation numérique. Ainsi, les spécialistes des Sciences de la vie les utilisent pour explorer le vivant (recherche fondamentale sur les maladies et épidémies, développement de médicaments, essais cliniques numériques, jumeau numérique humain, médecine personnalisée, neurosciences, etc.).

De même, les spécialistes des Sciences de la terre ont largement recours aux technologies numériques pour étudier la terre, l'eau et l'air, afin de trouver de nouvelles ressources, prévoir et anticiper des phénomènes naturels, mieux gérer l'environnement, prédire les évolutions climatiques, faciliter le développement durable, etc.

Et l'on retrouve aussi maintenant les technologies numériques dans les Sciences des matériaux, la culture, le patrimoine, la gestion des territoires, etc., avec de plus en plus d'applications grand public grâce à la 3D et aux technologies de réalité augmentée ou virtuelle.

Des évolutions présentes et à venir que les meilleurs experts internationaux vous présentent dans les pages qui suivent. ■

BIOLOGIE NUMÉRIQUE POUR EXPLORER ET PRÉDIRE LE VIVANT

Marjorie Domergue

Cheffe de projet du métaprogramme DIGIT-BIO, INRAE

Hervé Monod

Directeur du métaprogramme DIGIT-BIO, INRAE

Carole Caranta

Directrice générale déléguée Science et Innovation et co-directrice de DIGIT-BIO, INRAE

Ces dernières années, nous assistons à une véritable explosion des données en biologie, grâce aux nouvelles technologies d'acquisition des données et à l'essor de la science ouverte.

La recherche en sciences du vivant est profondément impactée par ce déluge de données, associé aux avancées des outils pour les analyser (modélisation, intelligence artificielle, calcul intensif). La biologie numérique nous permet aujourd'hui de développer des approches plus intégratives des systèmes biologiques, pour les explorer dans toute leur complexité, en intégrant différents niveaux d'échelles spatiales.

Dans ce contexte, le métaprogramme DIGIT-BIO¹, lancé par INRAE en 2021 a pour objectif de soutenir des recherches interdisciplinaires à l'interface entre sciences formelles et sciences du vivant pour mieux comprendre le comportement de systèmes biologiques, en prédire l'évolution dans des environnements changeants et faciliter leur pilotage.

Les projets ci-dessous illustrent comment le numérique et la science des données imprègnent les recherches actuelles en sciences du vivant. Les futurs développements feront encore plus largement appel aux méthodologies utilisant l'IA et l'IA générative, ainsi qu'aux jumeaux numériques², qui offrent un nouveau cadre pour améliorer notre capacité à agir sur des systèmes, en mobilisant leurs représentations numériques.

L'IA AU SERVICE DE LA SÉLECTION GÉNÉTIQUE PORCINE

Le projet OBAMA associe intelligence artificielle et génomique pour améliorer la sélection porcine, grâce à la prédiction de l'effet des mutations génétiques sur des caractères clés comme la croissance, le comportement ou la qualité de la viande.

La génomique a révolutionné la sélection génétique animale. En séquençant l'ADN d'un grand nombre d'animaux, on peut identifier des milliers de variants génétiques associés à des traits d'intérêt, tels que la fertilité ou le comportement. Cependant la majorité de ces variants sont situés dans des régions non codantes de l'ADN, rendant difficile la compréhension de leur effet réel.

Pour relever ce défi, deux chercheurs du centre INRAE d'Occitanie Toulouse - Julie Demars, généticienne et Raphaël Mourad, expert en Deep learning - ont initié le projet interdisciplinaire OBAMA, avec un double objectif :

- Développer de nouvelles approches d'apprentissage profond capables d'identifier avec précision les variants causaux de caractères d'intérêt, grâce à l'augmentation par orthologie de données issues du séquençage de génomes de plusieurs espèces
- Valider expérimentalement la prédiction de l'impact phénotypique des variants obtenus grâce à ces modèles, pour un trait d'intérêt donné.

1. <https://digitbio.hub.inrae.fr/>

2. <https://digitbio.hub.inrae.fr/thematiques/vers-des-jumeaux-numeriques>



© @REZOOMMarketing

Le projet s'appuie sur les données d'un programme pionnier de génétique porcine (PorcQTL), croisant des porcs de races Large White et Meishan. Les données recueillies incluent des traits classiques, comme le poids ou la composition de la viande, mais aussi des données comportementales, comme la docilité ou l'attention maternelle.

Grâce aux études d'association pangénomiques (GWAS), il est possible d'identifier des régions du génome statistiquement liées à ces caractères. Chaque région peut contenir des milliers de variants et seuls quelques-uns jouent un rôle causal.

C'est à cette étape qu'intervient l'apprentissage profond, qui va permettre de prédire l'impact de mutations, non pas directement sur le phénotype, mais sur certains phénomènes moléculaires : transcription des gènes, activité de la chromatine...

L'approche classique d'apprentissage supervisé est limitée par le faible volume de données fonctionnelles disponibles chez le porc. Pour lever ce verrou, OBAMA propose une stratégie d'augmentation de données par orthologie, inspirée des techniques utilisées en traitement d'images : les très nombreux génomes de mammifères déjà séquencés, mais non annotés, permettent d'enrichir massivement les jeux d'apprentissage, afin de développer des modèles plus robustes et plus précis.

À terme, le projet pourrait déboucher sur une stratégie inédite d'identification des mutations responsables de caractères complexes également chez plusieurs espèces d'élevage.

L'ÉPIGÉNÉTIQUE INTÉGRATIVE POUR PRÉDIRE LES CAPACITÉS ADAPTATIVES DES BIOAGRESSEURS

Dans un contexte de réduction d'utilisation d'intrants ou face aux effets du changement climatique, l'agriculture doit aujourd'hui faire face au développement de nombreux pathogènes. Comprendre et prédire les capacités d'adaptation de ces bioagresseurs à leur environnement est un enjeu crucial pour une lutte efficace, abordé par le projet EIPREDICT.

Les variations épigénétiques sont des modifications réversibles et héréditaires de l'expression des gènes, sans altération de la séquence de l'ADN. En réponse



© Shipher Wu (photograph) and Gee-way Lin (aphid provision),
National Taiwan University – PLoS Biology

à des contraintes environnementales sur de courtes périodes, comme des épisodes de fortes chaleurs, ces modifications sont un mécanisme adaptatif remarquable permettant aux organismes de produire de nouveaux phénotypes, favorisant leur survie et leur développement.

L'étude de ce code épigénétique s'appuie sur des techniques de séquençage à haut débit, qui génèrent des données massives, que les méthodes d'analyse actuelles peinent à exploiter complètement.

Le projet EPIPREDICT, associant deux chercheurs en génomique d'INRAE, Nadia Ponts (MycSA) et Gaël Le Trionnaire (IGEPP) à un chercheur en statistiques, David Causeur (IRMAR), propose de lever ce verrou.

L'ambition est de développer des méthodes statistiques et mathématiques innovantes pour identifier les éléments épigénétiques responsables des variations de l'expression des gènes sur deux modèles d'études : le puceron du pois et le champignon phytopathogène *Fusarium graminearum*.

Pour comprendre l'association entre épigénome et transcriptome, différentes méthodologies ont été testées. Les études comparatives démontrent la supériorité de modèles de mélange de régressions fonctionnelles, qui sont aussi performants que les méthodes d'apprentissage profond dans la prédiction de l'expression, et porteurs d'informations biologiques pertinentes sur les différents motifs d'association entre ouverture de la chromatine et expression des gènes cohabitant au sein d'un génome.

Ce projet permettra de construire de nouveaux modèles prédictifs, de comprendre plus finement des mécanismes épigénétiques, et de contribuer à la conception d'agro-écosystèmes résilients et durables. ■



© @Pexels Sarai Zuno

SANTÉ & BIOLOGIE : MÉDECINE DE PRÉCISION ET INTELLIGENCE ARTIFICIELLE

Claude P. Bertrand

Executive Vice-President Research & Development, Chief Scientific Officer, Servier

Les patients n'ont pas le temps d'attendre ! Pourtant, le processus de R&D en Santé est un processus long et coûteux car risqué et fortement réglementé.

Aujourd'hui, il faut 10 ans au minimum pour mettre un nouveau médicament sur le marché en tenant compte de l'ensemble du cycle Recherche & Développement (R&D). L'investissement nécessaire s'élève à environ 2 Md€, avec un taux d'attrition pouvant atteindre 90 % !

Chaque mois gagné, chaque année gagnée sont essentiels pour les patients qui attendent un traitement.

L'Intelligence Artificielle (IA) et les nouvelles technologies de la Data jouent un rôle essentiel dans la R&D pharmaceutique. Nous sommes convaincus qu'en combinant l'innovation scientifique et la puissance de la technologie, nous ouvrirons la voie à de nouveaux traitements révolutionnaires, en particulier pour les patients souffrant de maladies rares. Nous estimons que l'IA aurait le potentiel d'améliorer de 10 à 15 % la probabilité de succès dans le développement de médicaments et de réduire potentiellement le délai de mise sur le marché de deux à quatre ans.

AU DÉBUT ÉTAIT LA MÉDECINE DE PRÉCISION

La médecine de précision est une approche qui prend en compte les variations génétiques, environnementales et de mode de vie de groupes d'individus pour concevoir des stratégies de prévention, de diagnostic et de traitement adaptées. Contrairement à la médecine

traditionnelle, qui applique des traitements standardisés à tous les patients, la médecine de précision cherche à identifier les interventions les plus efficaces en fonction des caractéristiques propres à chaque groupe. À terme, on pourrait imaginer une évolution vers une médecine personnalisée ou un traitement correspondant à un patient.

Concrètement, la médecine de précision a déjà permis des avancées significatives à plusieurs étapes du cycle de vie du médicament :

- Diagnostic précoce grâce à l'utilisation de tests génétiques pour détecter des maladies à un stade précoce, les « companion diagnostics »
- Stratification des patients : la capacité à isoler des sous-groupes de patients en fonction de leurs caractéristiques génétiques et cliniques permet des diagnostics plus précis.
- Thérapies ciblées : en ciblant spécifiquement les mutations génétiques ou des anomalies biologiques chez un patient, on accroît significativement les probabilités d'efficacité du futur médicament...

DES APPLICATIONS ET AVANTAGES DÉCOUPLÉS PAR L'USAGE DES DATAS ET DE L'IA

Ces technologies jouent le rôle d'un accélérateur sans précédent pour analyser les composantes de la médecine de précision, transformant en profondeur les façons de faire et surtout de réussir en R&D.

Deux cas d'usages essentiels :

Bases de données de santé : Utiliser et partager de vastes ensembles de données provenant de dossiers médicaux électroniques, d'études cliniques et de recherches pour identifier des tendances et des corrélations.

Algorithmes d'IA : Appliquer des algorithmes d'apprentissage automatique pour analyser les données et fournir des recommandations personnalisées.

Grâce aux avancées technologiques et à l'analyse des données toujours plus performantes, la médecine de précision promet de transformer la manière dont les maladies sont diagnostiquées, traitées et prévenues.

L'IA est intégrée dans de nombreux aspects de la médecine de précision, de la découverte de médicaments à la personnalisation des traitements en passant par la pharmacovigilance post-commercialisation.

L'IA est essentielle à la médecine de précision car elle permet de traiter et d'analyser des données complexes et massives, d'identifier des biomarqueurs et des cibles thérapeutiques, de personnaliser les traitements, de prédire les réponses au traitement, d'accélérer la recherche et le développement, d'améliorer la précision diagnostique et de faciliter la médecine préventive. Grâce à l'IA, la médecine de précision peut offrir des soins plus adaptés, efficaces et

sûrs, transformant ainsi la manière dont les maladies sont diagnostiquées, traitées et prévenues.

Ainsi l'IA transforme la R&D en Santé en rendant les processus plus rapides, plus précis et plus personnalisés, ce qui peut conduire à des découvertes médicales révolutionnaires et à des traitements plus efficaces et sûrs.

IA ET DATA AU SERVICE DU PATIENT

Personnalisation des traitements :

- Utilisation des données pour administrer des thérapies personnalisées et adaptées aux besoins individuels des patients.
- Analyse des données génétiques et cliniques pour formuler des plans de traitement spécifiques à des populations spécifiques de patients.

Suivi et Prévention :

- Mise en place de systèmes de suivi des patients basés sur l'exploitation des données en temps réel pour anticiper les problèmes de santé.
- Développement d'algorithmes prédictifs pour détecter précocement des maladies ou des rechutes.

Système de soins centrés sur le patient : Utilisation des données pour améliorer la communication et la coordination entre les différents professionnels de santé impliqués dans le parcours de soins du patient. ■

CHIFFRES CLÉS

Investissements en IA dans la santé :

- En 2011, les investissements mondiaux en IA pour la santé étaient relativement modestes, avec moins de 100 millions de dollars investis.
- En 2021, ces investissements ont explosé, atteignant environ 6,6 milliards de dollars, selon les rapports de CB Insights et d'autres sources de capital-risque. Une part importante de ces investissements est consacrée à des applications en oncologie.

Nombre de publications scientifiques :

- Le nombre de publications scientifiques sur l'utilisation de l'IA en oncologie a augmenté de manière

exponentielle. En 2011, il y avait environ 200 publications sur ce sujet.

- En 2021, ce nombre avait dépassé les 2 500 publications par an, selon des bases de données comme PubMed et Google Scholar.

Essais cliniques utilisant l'IA :

- En 2011, il y avait très peu d'essais cliniques utilisant l'IA en oncologie, avec moins de 10 essais enregistrés.
- En 2021, le nombre d'essais cliniques utilisant l'IA pour le diagnostic, le traitement et la gestion du cancer avait augmenté à plus de 300, selon ClinicalTrials.gov.

APPORTS DU HPC ET DE L'IA À LA SANTÉ NUMÉRIQUE : ENJEUX ET DÉFIS

Yves Vandenbrouck

Directeur de Recherche, Maison de la Simulation, CEA

Selon l'Agence du numérique en santé (ANS), 30 % des données mondiales concernent la santé, et ce volume double tous les 73 jours. L'initiative UK Biobank, avec ses données génomiques et d'imagerie de plus de 500 000 individus, représente 30 Petabytes. En 2024, les investissements européens en santé numérique ont augmenté de 17 %, pour un total mondial de 24,5 Md\$, dont près de la moitié aux États-Unis. L'exploitation de ces données par l'IA ou les jumeaux numériques ouvre des perspectives : découverte de molécules, identification de biomarqueurs, essais cliniques virtuels, diagnostic, médecine personnalisée, prévention, optimisation des soins. Les principaux défis concernent la gestion de données massives et hétérogènes, la réduction des délais de recherche, et la personnalisation des traitements, nécessitant des capacités d'analyse avancées.

COMBINER IA, HPC ET BIG DATA POUR TRANSFORMER LA RECHERCHE BIOMÉDICALE ET LA MÉDECINE DU FUTUR

Les avancées du Deep Learning, combinées au calcul intensif et au big data, ont révolutionné la modélisation 3D des protéines. Sur 200 millions connues, seules 20 % ont une structure tridimensionnelle identifiée. Depuis des années, des approches *in silico* visent à prédire ces structures, comprendre les pathologies et concevoir de nouveaux traitements. Une rupture majeure vient des travaux de David Baker (conception « de novo » de protéines) et de l'équipe de DeepMind (Demis Hassabis,

John Jumper), récompensés par le prix Nobel de chimie 2024. Leurs résultats, avec AlphaFold, ont fait bondir la précision de prédiction de 40 à plus de 90 %, grâce à une IA à base de mécanismes d'attention, entraînée sur plus de 2 milliards de séquences et déployée sur 128 TPU V3 optimisés pour les réseaux neuronaux (cf. Figure 1).

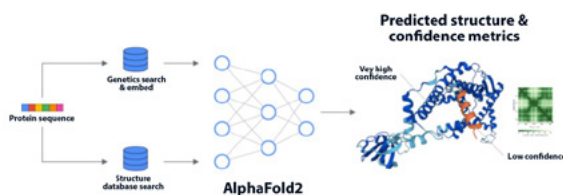


Figure 1. Diagramme simplifié du flux d'inférence AlphaFold2 (source ©EMBL-EBI).

Malgré des défis persistants (comme la reconnaissance antigène-anticorps), le succès de l'IA en santé repose sur l'alliance de compétences multidisciplinaires et d'infrastructures de calcul puissantes. Ces avancées ouvrent de nouvelles perspectives, notamment pour découvrir des traitements. En 2025, l'IA générative a permis d'identifier une cible de la fibrose pulmonaire et de créer une molécule inhibitrice prometteuse, testée avec succès en phase II par Insilico Medicine — une première mondiale. Compte tenu des coûts (50 à 100 M€) et du faible taux de succès (10 à 20 %) pour développer un candidat, ces résultats illustrent le potentiel de l'IA. Des start-up comme IKTOS ou InstaDeep exploitent le cloud et l'IA pour explorer d'immenses espaces chimiques,

automatiser la synthèse et planifier la robotisation, équivalente à la productivité de 30 chimistes. Par ailleurs, les modèles de fondation (FM) émergent en biologie, capables d'exploiter de vastes volumes de données non étiquetées. Un projet national du CEA vise à développer un FM pour analyser les repliements du cortex humain, à partir d'images cérébrales de plus de 100 000 individus (UK Biobank), afin d'identifier des marqueurs précoces de troubles psychiatriques. Toutefois, malgré ces promesses, le niveau de maturité de l'IA reste variable selon les domaines et nécessite un encadrement éthique dès la conception (« ethics by design »).

NIVEAUX DE MATURITÉ DES CAS D'USAGE ET DE L'IA EN SANTÉ

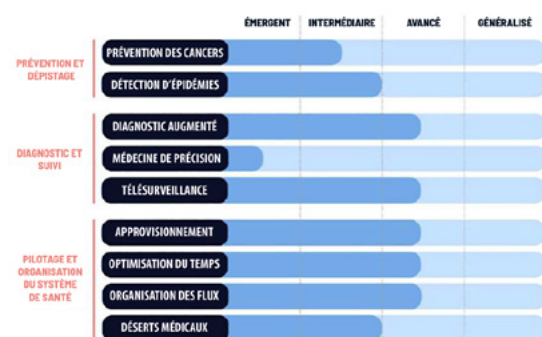


Figure 2. Degré de maturité des cas d'usage de l'IA en santé
(© HealthCare Data Institute)

STRUCTURATION DE L'ÉCOSYSTÈME DE LA SANTÉ NUMÉRIQUE EN FRANCE : PROGRÈS ET VEROUS À LEVER

Lancée en janvier 2021 dans le cadre de France 2030, la stratégie d'accélération en santé numérique vise à structurer un écosystème français compétitif à l'échelle mondiale. Elle s'articule autour de la formation (80 M€), de la recherche via le PEPR « Santé numérique » (40 M€), et de l'évaluation des dispositifs médicaux numériques (20 M€). Un enjeu clé réside dans le déploiement d'infrastructures numériques puissantes, sécurisées et conformes aux exigences réglementaires (RGPD, CNIL, SecNumCloud). Un appel à projets de 75 M€ a ainsi été lancé en 2022 pour renforcer les entrepôts de données hospitaliers, avec

16 établissements lauréats visant à constituer un réseau national de partage sécurisé des données de santé. Dans un marché en forte expansion, les GAFAM, comme AWS (UK Biobank) et Microsoft Azure (SNDS), s'implantent, suscitant des débats sur la souveraineté des données. La plate-forme nationale de données de santé (ou HDH) devra proposer un hébergement souverain certifié d'ici 2026. Ce contexte incite les acteurs français et européens à adapter leur offre pour garantir sécurité et accès aux données. L'interopérabilité reste un chantier prioritaire pour réutiliser efficacement des données souvent éparpillées, hétérogènes et peu documentées. Lever les freins à l'accès (lourdeur administrative, silos) et promouvoir une culture de partage entre producteurs et utilisateurs est crucial. Le rapport Marchand-Arvier (2024) appelle à un pilotage interministériel renforcé et à une refonte de la gouvernance du HDH. La publication du règlement européen sur l'Espace européen des données de santé (EEDS), imposant un guichet (ORAD) dans chaque pays d'ici 2026, devrait accélérer la coordination et l'accès rationalisé aux données de santé.

CONSTRUIRE L'ÉCOSYSTÈME DE LA SANTÉ NUMÉRIQUE

L'essor de la santé numérique en France repose sur la capacité à maîtriser et combiner HPC, big data et IA, dans un contexte international très concurrentiel. Ce changement profond, à la fois culturel, technologique et organisationnel, exige des investissements dans la formation (data science, cybersécurité, etc.) et dans des infrastructures souveraines et sécurisées, avec des architectures logicielles IA fiables, auditables et interprétables. Outre les initiatives nationales, des programmes comme NumPex (Exascale/IA) ou les projets européens (EuroHPC, EBrains, EEDS), co-pilotés par la France, offrent des opportunités pour mutualiser compétences et ressources. L'atelier « HPC/IA pour la santé numérique » du Forum Teratec 2025 a illustré la nécessité de coopérations entre industriels, chercheurs, cliniciens et producteurs de données. Si des synergies restent à renforcer, les talents et projets sont là. Il est donc essentiel que tous les acteurs — publics et privés — s'organisent pour construire un écosystème collaboratif au service de la recherche biomédicale et de l'amélioration du parcours patient. ■

NEUROSPIN : UN MARIAGE DE LA PHYSIQUE ET DU NUMÉRIQUE POUR SERVIR LES NEUROSCIENCES

Jean-François Mangin

Directeur de l'UMR Baobab à Neurospin, CEA

Le monde de la physique a compris depuis longtemps la nécessité de disposer de très grands instruments pour répondre à certaines questions fondamentales. Celui des sciences de la vie n'a pas encore connu une telle transition de phase. Neurospin a été fondé par le CEA il y a une vingtaine d'années pour tenter d'initier un tel processus, avec au cœur de son projet une IRM hors-norme appelée Iseult. Son aimant de 11,7 T a été conçu par les physiciens qui travaillent usuellement pour le domaine de la physique des hautes énergies. Le premier enseignement de ce projet c'est que le temps de la physique n'est pas celui

de la biologie : à la surprise des instigateurs de ce projet, il aura fallu presque vingt ans pour réaliser les premières images in vivo. Ce jalon a marqué le passage de témoin aux physiciens de Neurospin et à leur expertise de l'imagerie. L'image de la Fig. 1, qui dépeint la morphologie très plissée du cervelet, illustre le potentiel sans pareil de Iseult. Sans la résolution hors-norme de cette image (300 microns isotropes), il est impossible de visualiser le côté presque fractal de cette anatomie. C'est sur le plan de l'imagerie fonctionnelle que se focalisent maintenant les efforts pour gagner en résolution spatiale. Le cortex humain est organisé en couches

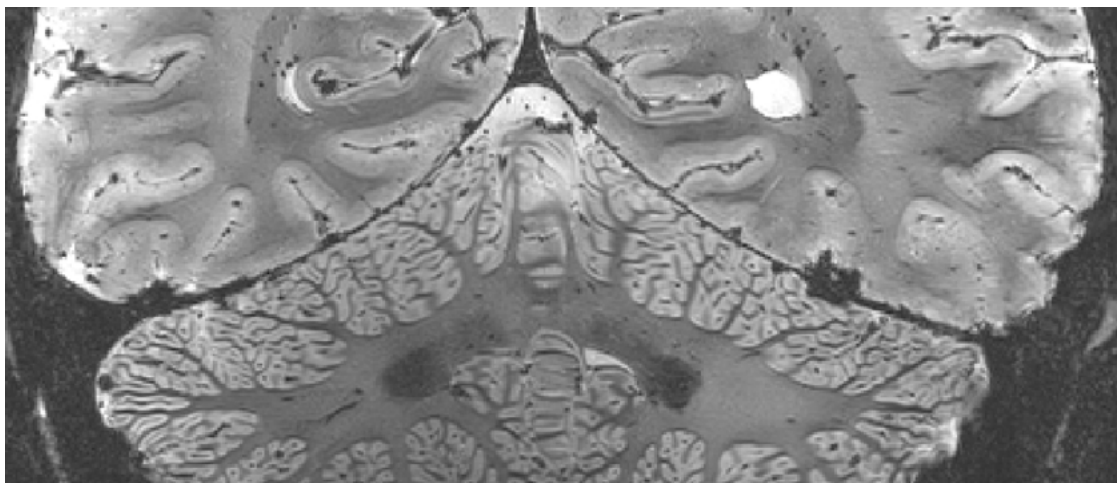


Fig. 1: Cervelet Humain imageé in vivo à 11,7T (300 microns isotrope) - © Courtesy, N. Boulant & A. Vignaud

et en colonnes de moins d'un millimètre d'épaisseur. Zoomer à une résolution permettant de distinguer les signaux issus de ces composants élémentaires est fort de promesses.

DE LA PHYSIQUE À L'ALGORITHMIQUE

D'autres corps de métier sont maintenant sollicités pour poursuivre l'aventure qui conduira aux premières découvertes en neurosciences : les experts de l'analyse de données qui développeront les algorithmes dédiés aux spécificités des images d'Iselt, et les chercheurs en neurosciences cognitive ou clinique qui vont concevoir les premières expériences. Dans la mesure où l'imagerie IRM est complètement digitale, le numérique va désormais jouer un rôle prépondérant aux côtés de la physique, à tous les niveaux de la chaîne d'analyse.

L'IRM est une machine qui permet de lire la « transformée de Fourier » de l'objet qu'on enfouit dans son tunnel. Cette lecture sur un mode cartésien peut prendre un temps considérable. Le groupe de recherche MIND (CEA-INRIA) utilise l'IA pour inventer de nouvelles trajectoires qui permettent d'optimiser le nombre de points de lecture, en combinant une vision de type « compress sensing » et une modélisation des contraintes induites par l'électronique de l'imageur. Cette stratégie va minimiser les temps d'acquisition.

La modélisation numérique s'impose également lorsqu'il s'agit de combiner un grand nombre d'images brutes pour révéler une structure complexe. L'équipe Ginkgo (CEA) a ainsi établi un record mondial en ce qui concerne la cartographie des connexions cérébrales. En accumulant 5 000 heures d'acquisitions post mortem sur un aimant 11,7T dédié aux rongeurs, après découpage du cerveau humain en cube, ils ont atteint une résolution de 200 microns isotrope. L'image du câblage calculée découle d'un processus d'optimisation massif, réalisé au TGCC, pour interpréter les perturbations des mouvements aléatoires de l'eau induites par les membranes des axones. Le modèle sous-jacent s'inspire des verres de spin de la physique statistique.

IMAGERIE ANALYSÉE PAR L'IA

C'est enfin au niveau de l'analyse de très grandes bases de données que le numérique prend définitivement son envol. Les évolutions les plus récentes de la science ouverte permettent aujourd'hui l'analyse de plus de 100 000 IRM de cerveaux, comme par exemple dans le cadre du projet Champollion du programme de recherche à risque Audace du CEA, qui vise à faire émerger une IA capable de décrypter le langage caché derrière les motifs de plissement du cortex de chaque individu (cf. Fig. 2). ■

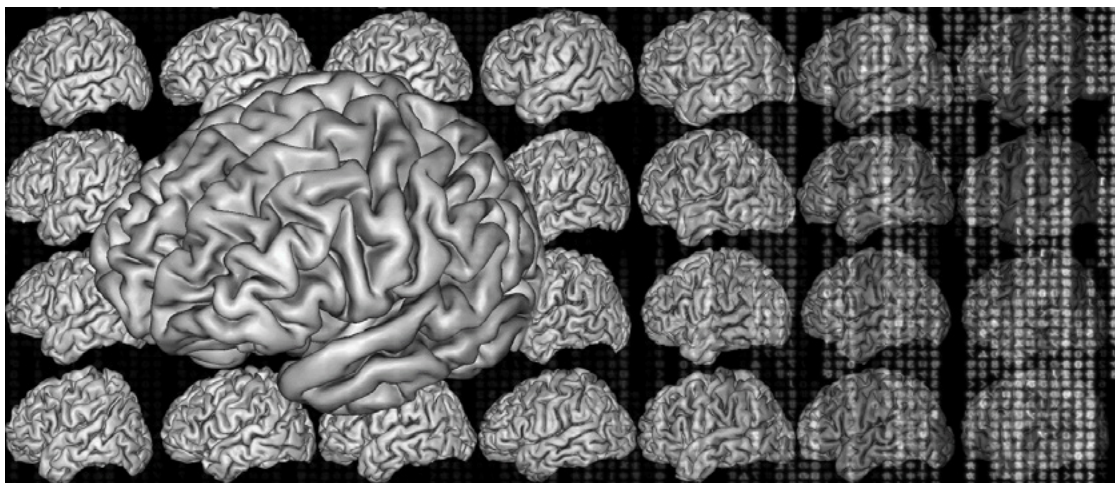


Fig. 2 : une fraction de la pierre de Rosette de 100000 cerveaux utilisée par l'IA Champollion pour décrypter les « idéogrammes du cortex »

LE JUMENT VIRTUEL HUMAIN : UNE AVANCÉE CRUCIALE POUR LA SCIENCE

Steven Levine,

Directeur Virtual Human Modeling, Dassault Systèmes

Débutée voici une quinzaine d'années, la modélisation numérique des organes humains s'est accélérée lors de la pandémie pour aboutir aujourd'hui à un véritable Jumeau Virtuel Humain (JVH), utilisable aussi bien en recherche qu'en clinique.

Débutés par la modélisation du cœur, Living Heart Project, ces travaux se sont étendus à d'autres organes (cerveau, foie, poumons, reins, pancréas...). Grâce à la plate-forme technique d'échange collaboratif développée par Dassault Systèmes, qui est utilisée par des milliers d'experts pour partager leurs recherches, et grâce aux financements publics, on dispose de modèles multi-organes et multi-échelles basés sur les données : paramétriques pour refléter la variabilité humaine ; très précis, on est passé de quelques millions à des centaines de millions de degrés de liberté ; utilisables en clinique car validés par les autorités réglementaires.

DES MODÈLES VALIDÉS PAR LA FDA, DES FONCTIONNALITÉS QUI EXPLOSENT

La précision des modèles dépend de l'organe et du contexte d'usage, mais deux critères clés guident leur validation : l'acceptation réglementaire et l'usage clinique. Une étude de 5 ans menée avec la FDA, a validé l'utilisation du jumeau du cœur comme substitut de population dans des essais cliniques in-silico. Plusieurs dispositifs médicaux ont ainsi obtenu leur agrément.

L'usage clinique des JVH représente la dernière étape. Il s'est étendu, par exemple pour guider des interventions chirurgicales complexes en pédiatrie. L'expérience montre que les chirurgiens disposant de JVH obtiennent de meilleurs résultats que les autres.

Les premiers jumeaux virtuels se concentraient sur quelques phénomènes bien compris pour valider la technologie. Aujourd'hui, leur champ d'action est uniquement

limité par la disponibilité des données. En fusionnant la physiologie mécanistique avec les données cliniques, les objets connectés, les données génomiques ou multi-omiques, un JVH peut prédire des résultats sur l'ensemble des activités humaines.

L'IMPACT DES TECHNOLOGIES NUMÉRIQUES

Les progrès technologiques continus (Simulation, IA, HPC, Quantique...) sont le moteur qui distingue les JVH par rapport aux expérimentations animales ou tests cliniques pour prédire la réponse humaine. Ainsi les modèles de cœur s'adaptent instantanément à n'importe quelle morphologie cardiaque. De même, le Living Liver Model combine la modélisation pharmacocinétique avec celle des fluides et des structures, couvrant ainsi plusieurs ordres de grandeur temporels et spatiaux.

La puissance de calcul disponible dans le cloud permet aux entreprises et aux hôpitaux d'être plus réactifs en exécutant des essais cliniques in-silico en parallèle sur des milliers de patients virtuels.

Enfin, l'IA crée rapidement des modèles spécifiques à une pathologie ou à un patient, avec moins de risque d'erreur que l'humain. De même, les premiers travaux menés avec des moteurs quantiques montrent qu'ils sont particulièrement adaptés à ce type d'applications.

LE HARDWARE-IN-THE-LOOP ARRIVE

On utilise aussi ces JVH pour tester des matériels en cours de développement. À terme on peut envisager des circuits ASIC spécialisés "Heart-on-Chip", mettant à jour en temps réels des champs électromécaniques, en analysant les données d'ECG, de pression ou de déformation issues de capteurs, pour piloter de nouveaux dispositifs médicaux intelligents.

Toutefois si le potentiel du JVH semble infini, il est limité par la connaissance encore parcellaire du fonctionnement du corps humain. La qualité et la précision de la mesure des données physiologiques des patients alimentant les modèles seront aussi cruciales. Enfin, passer d'une médecine basée sur l'observation à une médecine prédictive demande un changement profond de paradigme que la formation médicale devra anticiper. Une mutation en cours qu'il faut accélérer. ■

NUMÉRISER LE PATRIMOINE : DE LA PIERRE AU JUMEAU NUMÉRIQUE

Florian Moreno

Directeur R&D & Pôle Valorisation AGP

Voilà maintenant trois décennies qu'Art Graphique & Patrimoine (AGP) explore une idée simple, mais puissante : faire du numérique un allié au service de la mémoire des lieux.

Depuis notre création, nous avons vu le numérique transformer nos métiers. Au départ, il s'agissait d'aider à comprendre, documenter, dessiner ce qui existe — pour le restaurer, le protéger, parfois simplement pour ne pas l'oublier. Aujourd'hui, nous valorisons et rendons accessibles les sites via expériences immersives et doubles numériques interactifs. Et derrière tout cela : des relevés précis ; des modélisations lourdes ; des téraoctets de données à dompter. C'est tout un écosystème en mouvement, et AGP y tient sa place, avec exigence et pragmatisme.

UN PARCOURS QUI COMMENCE SUR LE CHANTIER

Quand Gaël Hamon a fondé AGP dans les années 1990, il venait de la pierre, au sens propre. Tailleur de pierre, il connaissait les monuments, les gestes du bâti, les contraintes du terrain. Mais il voyait aussi poindre quelque chose de nouveau : l'informatique qui allait, tôt ou tard, bouleverser la manière dont on regarde, mesure et pense le patrimoine.

Dès le départ, l'objectif était clair, produire des relevés rigoureux, fiables, exploitables pour les architectes du patrimoine et les entreprises spécialisées. Le numérique était un outil, pas un gadget. Il fallait l'apprendre avec la même exigence que le compas ou le tire-ligne.

Nous travaillions alors surtout sur le patrimoine bâti : façades, voûtes, portails sculptés. Mais vers 2005, nous avons élargi notre spectre. Les œuvres d'art muséales (statues, peintures, objets décoratifs) sont devenues un champ d'action à part entière. Là aussi, le besoin de documenter, conserver et valoriser était fort. Et les outils ont suivi.

LES OUTILS, UNE RÉVOLUTION SILENCIEUSE

Pendant des années, la priorité technique, c'était la précision. Chaque nouveau scanner promettait plus de détails et de densité. Et nous allions de l'avant. Jusqu'à ce qu'on atteigne, il y a une dizaine d'années, un plafond : au-delà d'un certain niveau de détail, l'œil humain ne perçoit plus la différence, et les usages non plus.

Depuis, vitesse, souplesse et compacité guident l'innovation. Là où il fallait deux opérateurs et quatre scanners pour relever une église, une seule personne



Relevé par drone du baptistère du Ve siècle de la cathédrale Saint-Léonce de Fréjus et le modèle numérique résultant. Doc AGP

NUMÉRISER LE PATRIMOINE : DE LA PIERRE AU JUMEAU NUMÉRIQUE

équipée d'un scanner dernier cri peut faire mieux, plus vite, et sans compromis sur la qualité. Les recalages deviennent semi-automatiques, les interfaces plus intuitives, et surtout, on peut croiser les approches : lasergrammétrie terrestre ; photogrammétrie au sol ou aérienne, grâce aux drones. Chaque monument est capté sous tous les angles, au millimètre près, avec souplesse d'intervention.

Mais tout cela entraîne une conséquence bien concrète : plus les outils deviennent performants, plus les fichiers deviennent lourds.

L'EFFET BOULE DE NEIGE DES DONNÉES

Entre 1994 et 2004, tous nos projets tenaient sur un seul disque dur. Aujourd'hui, il faut parfois plusieurs téraoctets pour stocker un chantier complet. Ce n'est pas qu'une anecdote technologique. C'est un défi quotidien.

Plus la captation est rapide et riche, plus les fichiers s'alourdissent. Cela exige infrastructure solide, archivage long terme, serveurs fiables, maintenance et sauvegardes. Un patrimoine numérique, en somme, qu'il faut préserver avec autant de soin que les pierres qu'il représente.

Mais la question n'est pas seulement de stocker. Il faut aussi traiter. Nettoyer les données, les assembler, les

optimiser, pour les rendre exploitables — que ce soit dans un logiciel métier ou dans une application de médiation. Et là, il faut faire preuve d'ingéniosité.

Depuis une dizaine d'années, nous avons intégré les logiques du rendu PBR (Physically Based Rendering). On ne se contente plus d'appliquer une simple texture. On travaille les « Normal Maps » qui permettent de simuler le relief d'une surface sans avoir recours à une géométrie complexe tout en feignant les reliefs, les effets de rugosité, de brillance, etc. Tout ça permet de produire des rendus bluffants, mais surtout légers. C'est un compromis subtil entre qualité perçue et poids des fichiers. Un travail invisible mais indispensable.

CE QUE LE CALCUL CHANGE (ET CHANGERA ENCORE)

Avec des volumes pareils, le traitement manuel n'est plus tenable. Le calcul haute performance (HPC) est devenu pour nous un partenaire discret, mais essentiel. Le recalage des nuages de points, la génération des maillages, le « Baking » des textures pour réduire le poids des modèles : tout cela se fait désormais en parallèle, sur des serveurs équipés de cartes graphiques puissantes. Et sans cela, certains projets resteraient bloqués à l'étape du traitement.



JumEAU numérique qui sert à la restauration en cours du salon de Diane au Château de Versailles. Doc AGP



Nuage de points du relevé de l'Arc de Triomphe d'Orange. Doc AGP

Ce n'est qu'un début. L'Intelligence Artificielle commence à prendre le relais sur des tâches fastidieuses : réaliser un détourage de centaines d'images ; classifier des gigas de données ; identifier des formes ; proposer des alignements... Elle ne remplace pas l'expertise humaine, mais soulage les équipes, accélère les flux et ouvre des perspectives.

Au-delà de la reconnaissance d'image elle aidera, comme dans le médical, à diagnostiquer la pathologie d'un bâtiment ou d'une œuvre. Dans la médiation, elle permettra un médiateur virtuel répondant finement et de façon personnalisée aux visiteurs.

Nous entrons dans l'ère du jumeau numérique 2.0 : non plus seulement visualiser un monument, mais le suivre, le surveiller et simuler des scénarios. Un modèle 3D devient outil d'analyse et plateforme de médiation. Cela a été notamment le cas avec le jumeau numérique de Notre-Dame de Paris après incendie. Couplé à des capteurs physiques, il peut même réagir, signaler une détérioration, anticiper des besoins de restauration.

ET MAINTENANT ?

Aujourd'hui, AGP a rejoint le pôle Memorist, aux côtés de cinq entités visant un même objectif :

préserver, numériser, enrichir et transmettre le patrimoine. Ce rapprochement s'est accompagné d'un changement d'équipe dirigeante, mais l'ADN reste le même : rigueur ; innovation ; respect de la matière et goût du détail.

Demain, il s'agit d'inscrire le numérique dans le temps long du patrimoine. Les relevés ne seront plus de simples instantanés, mais des flux réguliers, documentés, mis à jour. Chaque monument ou œuvre pourra exister dans un espace numérique vivant et interrogeable.

Le patrimoine sera plus largement accessible. Un visiteur explorera une sculpture en réalité augmentée depuis chez lui ; un restaurateur testera virtuellement une intervention ; un conservateur suivra l'évolution d'un objet grâce à des relevés croisés et visualisations dynamiques.

Ce type d'approche, déjà courant en médecine, émerge dans le monde patrimonial. Et même si tout reste à inventer, le potentiel est immense : conservation préventive ; formation ; transmission ; ouverture au grand public.

Et derrière tout cela, AGP restera à sa place : faire le lien. Entre la main et la machine. Entre la matière et le modèle. Entre le passé qu'on reçoit et l'avenir qu'on prépare. ■

LE RÔLE ET L'APPORT DES TECHNOLOGIES HPC, IA ET QUANTIQUE FACE AUX DÉFIS CLIMATIQUES ET ENVIRONNEMENTAUX

Philippe Drobinski

Directeur de Recherche CNRS, Professeur à l'Ecole polytechnique
Laboratoire de Météorologie Dynamique – Institut Pierre Simon Laplace (LMD-IPSL)
Directeur du centre interdisciplinaire Energy4Climate (E4C)

Le rôle et l'apport des technologies HPC, IA et quantique face aux défis climatiques et environnementaux

Face à l'urgence climatique, les technologies de pointe telles que le calcul haute performance (HPC), l'intelligence artificielle (IA) et l'informatique quantique se révèlent être des outils essentiels pour comprendre, modéliser et atténuer les effets du changement climatique. Ces technologies permettent de traiter des volumes massifs de données, de simuler des scénarios complexes et de développer des solutions innovantes pour une planète plus durable.

LE CALCUL HAUTE PERFORMANCE (HPC) : PILIER DE LA MODÉLISATION CLIMATIQUE

Le HPC est au cœur de la modélisation climatique moderne. Les supercalculateurs permettent de simuler des phénomènes atmosphériques, océaniques et terrestres avec une précision sans précédent. Par exemple, les modèles de circulation générale (GCM) utilisent le HPC pour prévoir les tendances climatiques à long terme et évaluer les impacts potentiels des émissions de gaz à effet de serre [1]. Des institutions telles que le National Center for Atmospheric Research (NCAR) aux États-Unis, le European Center for Medium-range Weather Forecast (ECMWF) en Europe, Météo France et l'Institut Pierre Simon Laplace (IPSL) en France exploitent des supercalculateurs pour améliorer la précision des prévisions météorologiques et climatiques.

Ces simulations aident les décideurs à élaborer des politiques d'adaptation et de mitigation plus efficaces.

L'INTELLIGENCE ARTIFICIELLE (IA) : CATALYSEUR DE SOLUTIONS DURABLES

L'IA transforme la manière dont nous abordons les défis environnementaux. Elle est utilisée pour analyser de vastes ensembles de données climatiques, optimiser les systèmes énergétiques et surveiller la biodiversité. Par exemple, des algorithmes d'IA peuvent prédire les événements météorologiques extrêmes, aidant ainsi les communautés à se préparer et à réagir plus efficacement. Des initiatives comme Aardvark Weather, développée par l'Université de Cambridge, utilisent l'IA pour fournir des prévisions météorologiques précises dans des régions où les ressources en supercalculateurs sont limitées [2]. De plus, l'IA est employée pour surveiller la déforestation, suivre les émissions de méthane, comme le fait l'entreprise Kayros, et optimiser l'agriculture de précision, contribuant ainsi à la réduction des émissions de gaz à effet de serre.

L'INFORMATIQUE QUANTIQUE : UNE PROMESSE POUR L'AVENIR

Bien que l'informatique quantique en soit encore à ses débuts, elle offre un potentiel considérable pour résoudre des problèmes complexes liés au climat. Les ordinateurs quantiques pourraient simuler des réactions chimiques pour développer de nouveaux matériaux de capture du carbone ou optimiser les

réseaux énergétiques pour une meilleure intégration des énergies renouvelables [3]. Des entreprises comme PsiQuantum travaillent sur des architectures quantiques capables de simuler des processus chimiques complexes, ouvrant la voie à des avancées dans la capture et le stockage du carbone. Cependant, des défis techniques subsistent, notamment en matière de stabilité des qubits et de consommation énergétique des systèmes de refroidissement nécessaires au fonctionnement des ordinateurs quantiques.

VERS UNE UTILISATION RESPONSABLE ET DURABLE

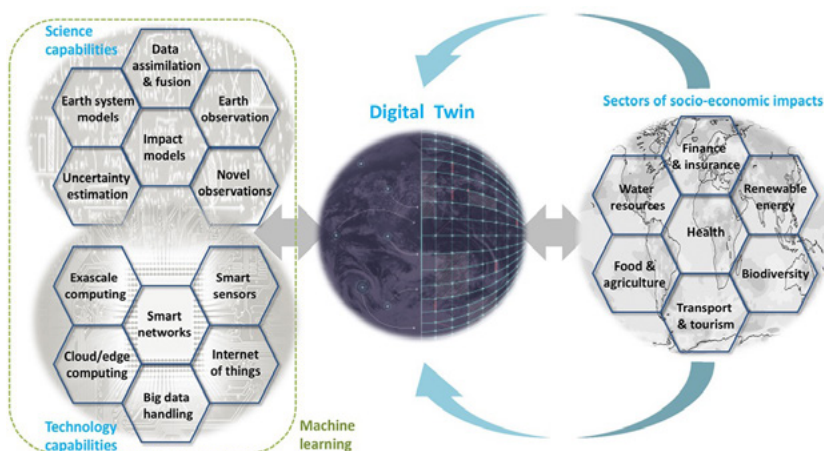
Si ces technologies offrent des solutions prometteuses, il est crucial de considérer leur propre impact environnemental. Les centres de données alimentant l'IA et le HPC consomment d'importantes quantités d'énergie et d'eau, contribuant aux émissions de gaz à effet de serre [4]. Il est donc essentiel d'adopter des pratiques durables, telles que l'utilisation d'énergies renouvelables, l'optimisation de l'efficacité énergétique et le recyclage des équipements électroniques. Des initiatives sont en cours pour rendre ces technologies plus écologiques. Par exemple, certaines entreprises développent des centres de données alimentés par des sources d'énergie renouvelables, tandis que d'autres travaillent sur des algorithmes d'IA plus efficaces, réduisant ainsi leur empreinte carbone. Des recherches spécifiques sur le numérique frugal sont également conduites dans des centres de recherche comme le centre Energy4Climate (E4C) en France.

CONCLUSION

Les technologies HPC, IA et quantique jouent un rôle crucial dans la lutte contre le changement climatique. Elles permettent de mieux comprendre les systèmes complexes de notre planète, de prévoir les impacts futurs et de développer des solutions innovantes pour un avenir durable. Cependant, leur déploiement doit être accompagné d'une réflexion sur leur propre impact environnemental, afin de garantir que ces outils servent véritablement la cause climatique. ■

Références

1. Kulkarni, P., Manoharan, S., & Gaddi, A. (2024). Advancing Climate Modeling through High-Performance Computing: Towards More Accurate and Efficient Simulations. EAI Endorsed Transactions on Energy Web, 11(4). <https://doi.org/10.4108/ew.7049publications.eai.eu>
2. Allen, A., Markou, S., Tebbutt, W. et al. End-to-end data-driven weather prediction. Nature 641, 1172-1179 (2025). <https://doi.org/10.1038/s41586-025-08897-0>
3. Giani, A., and Z. Goff-Eldredge (2022), How Quantum Computing Can Tackle Climate and Energy Challenges. Eos, 103, <https://doi.org/10.1029/2022EO220500>
4. MIT News. (2025). Explained: Generative AI's environmental impact. Retrieved from [https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117MIT News](https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117MIT%20News)
5. Bauer, P., Quintino, T. & Wedi, N. (2022) From the Scalability Programme to Destination Earth. ECMWF Newsletter, 171. <https://www.ecmwf.int/en/newsletter/171/earth-system-science/scalability-programme-destination-earth>



Credit : ECMWF [5]

Les jumeaux numériques du système Terre permettent aux secteurs socio-économiques d'accéder à des données de qualité optimale et de créer des outils permettant d'extraire des informations spécifiques à leurs applications. Un jumeau numérique s'appuie sur la convergence de compétences scientifiques et technologiques de pointe, basées sur des simulations et des observations, en exploitant l'ensemble des technologies numériques. Il s'appuie sur des méthodologies d'apprentissage automatique pour tous ses composants.

LE NUMÉRIQUE AU CŒUR DES GÉOSCIENCES POUR UN MONDE EN MUTATION

Michaël Chelle

Directeur de la stratégie numérique, Bureau des Recherches Géologiques et Minières (BRGM)

Romain Chassagne

Responsable Jumeaux numériques, Bureau des Recherches Géologiques et Minières (BRGM)

Simon Lopez

Responsable Calcul et simulation numérique, Bureau des Recherches Géologiques et Minières (BRGM)

Le sous-sol est devenu indispensable aux activités humaines dans un contexte de bouleversements climatiques, environnementaux, géopolitiques et numériques. Il fournit eau, énergies, métaux, solutions de stockage (CO_2 , chaleur, déchets), mais engendre aussi des risques en surface (séismes, mouvements de terrain, submersion, etc.). Or, ce milieu naturel est complexe et coûteux à explorer, au regard des différentes échelles spatio-temporelles (des minéraux micrométriques aux milliers de kilomètres carrés d'une plaque lithosphérique; des quelques secondes d'un séisme aux millions d'années des ères géologiques), de l'hétérogénéité de ses structures, de la non-linéarité des processus physico-chimiques qui y occurrent et des conditions extrêmes (température, pression, etc.) qui peuvent y régner. Pour gérer le sous-sol dans ce contexte, décideurs publics et privés ont besoin d'une « carte d'état-major » du sous-sol, à jour, en 3D et multiparamètres. Cet objectif ambitieux nécessite la transformation des géosciences par le numérique. C'est pourquoi, le BRGM en a fait un enjeu scientifique, visant à développer des géosciences quantitatives et prédictives, exploitant données et connaissances pour simuler les processus physiques et chimiques du sous-sol, tout en quantifiant les incertitudes. Cela permettra de mieux appuyer les décisions publiques, en produisant rapidement des données fiables et traçables, comme par exemple pour l'évacuation d'une zone menacée par une submersion marine.

Cela nécessite de décliner le numérique à travers différents prismes : les données, la simulation numérique, l'exploitation de l'intelligence artificielle (IA), et leur intégration dans des jumeaux numériques du sous-sol.

VALORISER LES DONNÉES

Depuis des décennies, le BRGM archive les données du sous-sol, à l'instar des services géologiques des autres pays. Un des enjeux actuels est de rendre ce patrimoine hétérogène et lacunaire plus accessible et utilisable. Pour cela, l'IA est prometteuse, notamment pour structurer et interpréter automatiquement ces données. Les mesures géophysiques, quant à elles, nécessitent des approches mathématiques et numériques pour résoudre des problèmes inverses et imager les entités géologiques profondes. Ces calculs, très gourmands en ressources CPU, s'appuient sur des codes parallélisés exploitant au mieux les nouvelles architectures.

SIMULATION NUMÉRIQUE DES PROCESSUS SOUTERRAINS

L'approche quantitative de nombreux enjeux liés au sous-sol repose sur la modélisation multiphysique des transferts et écoulements. Ces modèles peuvent être très complexes, selon le degré de couplage entre les phénomènes physiques et la nature plus ou moins hétérogène et discontinue des géométries géologiques, et donc consommateurs de calcul (Fig. 1). De plus ils servent à produire des cartes et des données

spécialisées, ce qui nécessite de multiplier les simulations tout comme la quantification des incertitudes. Le calcul parallèle a élargi leur usage, mais le temps de simulation reste un verrou, surtout en situation de crise. Pour le lever, il faut simplifier ces modèles (reduced order modeling), en réduisant leur complexité et/ou en développant des modèles moins gourmands en calcul (métamodélisation).

L'IA AFFRONTÉ LES GROS MODÈLES

Bien que limitée par la rareté des données du sous-sol, les usages de l'IA se développent rapidement en géoscience. Le machine learning est un excellent outil qui permet de simuler des phénomènes souterrains, malgré une connaissance partielle des mécanismes, et offre une exécution rapide en mode direct. Cette rapidité est aussi mise à profit pour développer des métamodèles à partir de simulations de « gros » modèles multi-physiques. Mais leur entraînement implique des simulations longues et exige donc un échantillonnage optimal de l'hyperespace des facteurs, tout en limitant l'incertitude.

L'IA générative, très consommatrice en GPU, est, elle, explorée, via des LLM et des PRAG, pour exploiter des

données non structurées (rapports, cartes anciennes, archives minières), mais aussi pour créer, via des GAN, des données synthétiques, compensant les difficultés d'observation directe.

JUMEAUX NUMÉRIQUES

Les jumeaux numériques émergent aussi pour le sous-sol, que ce soit pour les industries exploitant le sous-sol (géothermie) ou l'appui aux politiques publiques (gestion intégrée de l'eau, évaluation des risques de terrain). Véritables catalyseurs d'aide à la décision, ils reposent sur l'assimilation continue de flux de données monitoring dans des simulations sophistiquées. Ils ouvrent la voie à une compréhension dynamique et prédictive du système terrestre, tout en accroissant les besoins en calcul scientifique haute performance.

Cette transformation numérique des géosciences, indispensable pour répondre aux enjeux sociétaux liés au sous-sol, entraîne des besoins croissants en puissance de calcul, en CPU et parallélisme, mais aussi en GPU. Conscient des enjeux environnementaux et géopolitiques sur les ressources minérales, le BRGM vise toutefois à une utilisation raisonnée de ces ressources en calcul. ■

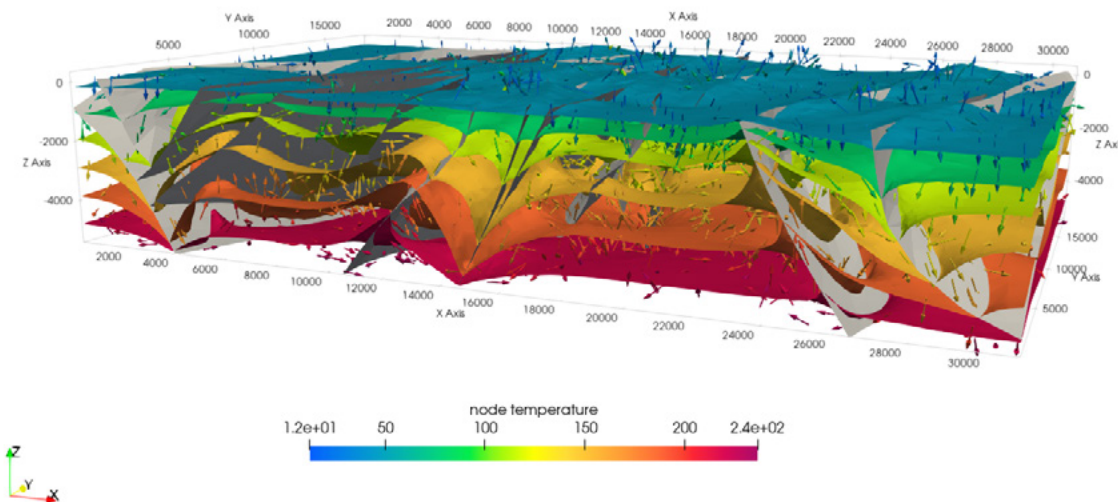


Figure 1: Champ de température dans un système géothermique du Fossé Rhénan, simulé par un modèle de dynamique hydrothermale, à base du code ComPASS (Armandine Les Landes A. et al, 2025, <https://doi.org/10.1016/j.cageo.2024.105752>)

EXTENSION ET HYBRIDATION DU HPC DANS L'AÉRONAUTIQUE

Thierry Chevalier

VP Chief Engineer Manufacturing & Process | Engineering R&D, Capgemini
Senior adviser complex system of systems modeling & simulation applied to climate transition

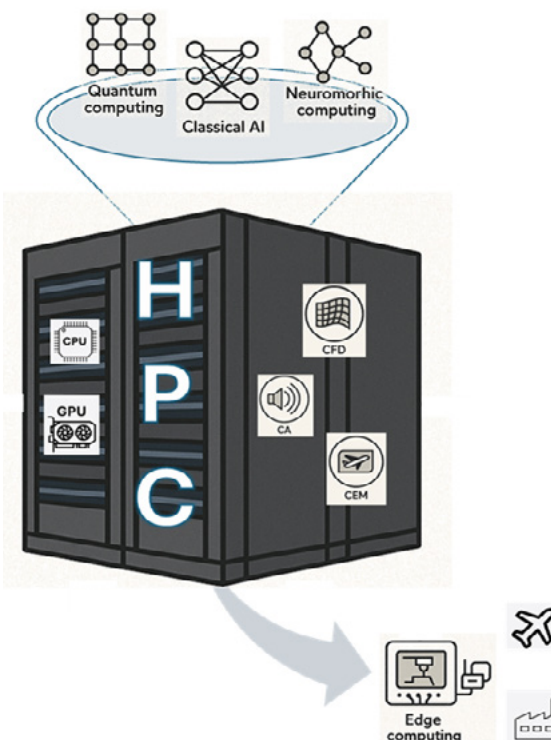
LES 10 DERNIÈRES ANNÉES

L'architecture HPC a significativement changé avec l'émergence de processeurs spécialisés (GPU, TPU, ARM parallèles, manycore Kalray...), permettant de contenir les coûts énergétiques des simulations. Cela a nécessité une refonte des codes, souvent en passant d'approches implicites à explicites, ou vers des méthodes plus consommatrices (LES, DES), devenues accessibles grâce aux calculateurs Exascale.

Les disciplines classiquement utilisatrices de HPC que sont l'aérodynamique, l'acoustique, l'électromagnétisme ou la résistance des matériaux ont tiré parti de ces avancées pour réduire voire supplanter les essais physiques, voire contribuer directement ou indirectement à la certification. L'optimisation multiphysique intégrée s'est développée, facilitant l'analyse couplée de phénomènes complexes (ex. analyses multidisciplinaires couplées aérodynamique/structure/acoustique/thermique d'installations motrices mât et réacteur).

Le calcul quantique, encore marginal, est exploré via des approches hybrides HPC/quantique, notamment pour la détection de pannes. L'IA s'intègre largement aux algorithmes numériques : initialisation rapide par IA de champs de solution réduisant le temps de convergence, modèles de turbulence IA robustes et performants sur des configurations très variées en CFD.

Les progrès en stockage (ex. RAM, disques, systèmes de fichiers distribués) et en matériel ont permis d'élargir



les cas traités : maillages plus fins, volumes de données accrus, modèles ML plus complexes. Cette amélioration des performances d'I/O ainsi que l'intégration de l'intelligence artificielle (IA) et de l'apprentissage automatique (ML) avec les systèmes HPC a révolutionné le traitement et l'analyse des données. Cela a favorisé les jumeaux numériques, la maintenance prédictive. L'explosion du nombre de compagnies aériennes ayant basculé vers de la maintenance à l'heure de vol est très révélatrice de ce succès qui permet de garantir du « Zero Aircraft On ground ». L'optimisation des trajectoires a aussi permis des gains comparables à une modernisation de flotte (ex. EasyJet).

L'intégration du HPC dans le cloud a facilité l'accès sans investissements lourds à la simulation, pour les nombreuses start-up émergentes dans l'aéronautique.

Enfin, la généralisation des techniques HPC vers le edge a révolutionné les opérations (ex : traitements d'image embarqués flexibles et pilotés depuis le sol) et la fabrication (ex. impression 3D avec une optimisation initiale des paramètres/formes à imprimer et des modèles d'apprentissage de paramétrisation de l'impression, ensuite embarqués dans du edge proche de l'imprimante, pour contrôler en temps réel les paramètres).

LES 10 PROCHAINES

Le calcul neuromorphique (ex. Tesla Dojo) améliore l'efficacité énergétique et permet l'apprentissage à partir de jeux de données réduits, grâce à sa capacité native à traiter les séquences temporelles. Cette approche devrait bénéficier à la maintenance préventive pour des modèles plus efficaces et évolutifs (volume de données d'apprentissage plus réduit pour techno récentes). Elle est éclipsée par le succès LLMs qui transforme la génération et l'exploitation de la documentation technique, le support, et même l'analyse de résultats de simulation.

Le même algorithme au cœur des LLM's (Google « Transformer ») ouvre aussi des perspectives au-delà du langage pour la gestion d'incertitudes et des marges (critiques pour des produits complexes comme les

avions), ou l'organisation des flux de données (ex. sur un bus avionique). La simulation numérique commence de plus à améliorer la structure des réseaux neuronaux sur lesquels il s'appuie (optim topologique ou catégorielle du réseau pendant sa création pour générer des réseaux « parcimonieux »), et rend possible à iso-performance des modèles allégés de plusieurs ordres de grandeur capable d'être utilisés en edge/embarqué.

L'intégration quantique-classique progresse, notamment via les qbits logiques, avec des applications en chimie quantique pour l'étude de nouveaux matériaux ou carburants alternatifs, enjeu clé pour une aviation post-kérosène.

L'augmentation de la puissance de calcul ouvre la voie à des simulations et optimisations quasi-temps réel, réduisant les cycles d'exploration et de conception - crucial à l'heure où on pense à bouleverser l'architecture classique « tube & wing - kérosène ».

L'efficacité énergétique croissante du HPC devrait continuer à compenser l'augmentation de ses usages, dans un contexte où les frontières entre simulation et IA s'estompent. Les approches hybrides deviennent la norme (ex. workflows partiellement pilotés par IA, mélanges de modèles de simulation et de ML...). À noter les capacités de l'IA à analyser les codes, proposer des portages technologiques, voire à assister ou automatiser la configuration et la maintenance des chaînes de simulation qui peut grandement faciliter l'évolution et la maintenance du large corpus existant de simulation numérique en aéronautique, à l'heure où les architectures matérielles évoluent rapidement.

Enfin, le HPC embarqué se généralise, dans les avions comme dans les usines. Associé à des modèles IA optimisés, il permet d'embarquer simulation, analyse et adaptation en temps réel au plus près des systèmes physiques. Cela ouvre la voie à des systèmes plus résilients, adaptatifs, et à une optimisation continue des performances en vol (ex. ajustement via LIDAR embarqué), avec une assistance accrue au pilotage. ■

TRANSFORMATION DIGITALE ET SIMULATION NUMERIQUE : L'ACCELERATEUR DES PROJETS VEHICULES

William Becamel

Expert Leader Simulation Numérique
Global Engineering Renault Group

Olivier Colmard

Directeur Transformation Digitale, CAE & PLM
Global Engineering Renault Group

Les défis du secteur automobile sont nombreux. Entre enjeux réglementaires, compétitivité internationale et concurrentielle accrue, transition énergétique et technologique (électrification des véhicules, aides à la conduite et systèmes connectés), la réponse de Renault Group est une ingénierie agile et rapide. C'est dans ce contexte que les développements de véhicules se déroulent en 2 ans dorénavant.

Pour atteindre ce niveau de performance, le digital est central. L'histoire n'a pas débuté aujourd'hui mais dès les années 2000 avec de forts investissements dans ces technologies. L'évolution des infrastructures

de calcul, en perpétuel progrès, a permis de réaliser des modèles de plus en plus précis et fiables (fig.1). L'exemple de la sécurité passive est particulièrement illustratif. Le progrès dans ce domaine a permis d'arriver à des niveaux de corrélations tels, qu'il est possible d'atteindre dès le premier crash les cibles de performance attendues.

SIMULATION DANS LE DÉVELOPPEMENT D'UN VÉHICULE : POURQUOI ET COMMENT

La généralisation de la simulation numérique est devenue incontournable dans un développement véhicule, c'est un attendu majeur des grands jalons

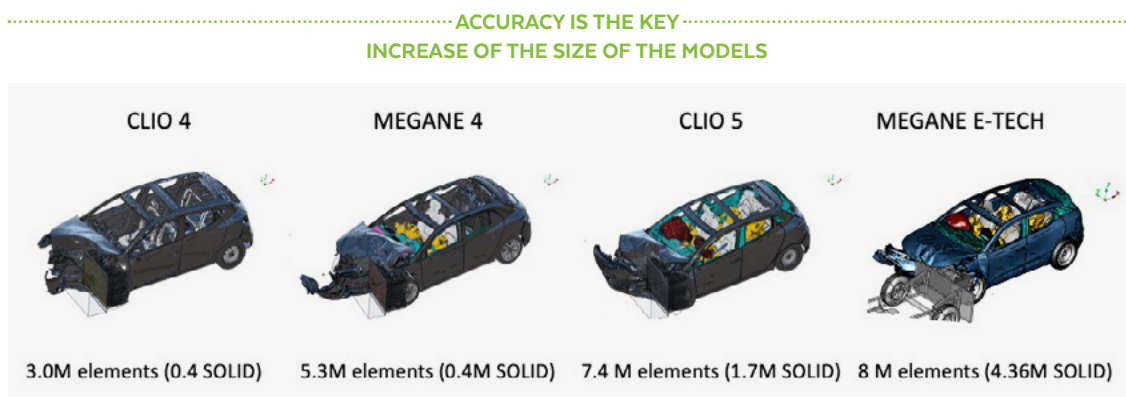


Fig.1: Accroissement de la taille des modèles entre différentes générations de véhicules. Progressivement, les moyens de retenus (airbag / ceintures) et les mannequins ont été par exemple intégrés. Cette évolution s'est faite sans changer le temps de restitution aux ingénieurs grâce aux évolutions d'infrastructure HPC et à l'optimisation des codes de calcul.

à franchir pour concevoir et industrialiser un produit. L'organisation a évolué pour répondre à ce besoin avec notamment la professionnalisation des métiers pour gérer les CAO en configuration, créer les maillages et les modèles, lancer et analyser les calculs. En seulement 3 semaines, nous sommes capables de produire le bilan de toutes les prestations structurantes d'un véhicule (aérodynamique, confort thermique, confort/comportement, endurance, acoustique, crash, ...).

Si des exigences ne sont pas respectées, des boucles itératives journalières sont menées par les ingénieurs pour résoudre les points durs et apporter les solutions au produit. Ce jumeau numérique du véhicule permet de concevoir rapidement, d'initier les phases de prototypes physiques avec un fort niveau de confiance et d'accélérer les validations finales. Les résultats sont à la hauteur de l'investissement car entre 2019 et 2022 par exemple, les coûts de validations ont été réduits de 53 % (fig.2) :

ACCÉLÉRER POUR RESTER DANS LA COURSE: LES LEVIERS DE LA SIMULATION NUMÉRIQUE

Cependant, nous ne pouvons pas nous arrêter là. La concurrence reste forte et la rupture nécessaire. Quelles sont les perspectives ?

1. Toujours améliorer notre couverture calcul (i.e. nombre d'activités de validation que nous pouvons simuler) et la maturité des simulations (i.e. leur précision)
2. accroître nos vitesses de simulation et modélisation. L'apport des GPU dans ce domaine est précieux selon les domaines physiques.
3. Coupler des moyens virtuels et physiques pour anticiper la mise au point des véhicules avant le début des phases industrielles. Nous parlons ici du continuum

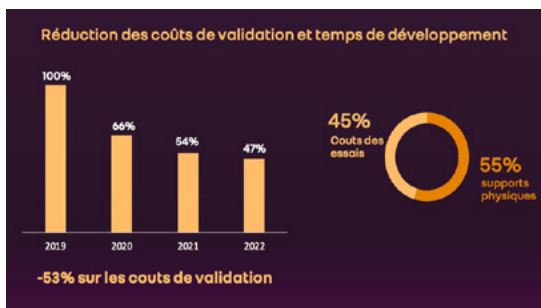
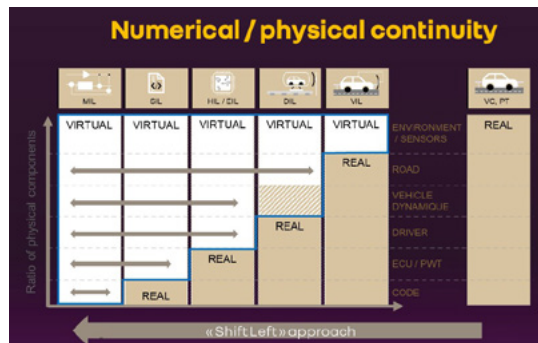


Fig.2: Évolution des coûts de validation Renault Group entre 2019 et 2022



des outils XiL (du Model in the Loop – MIL, Software in the Loop – SiL, Hardware in the Loop – HiL...)

4. Enfin, le dernier élément qui offre des opportunités de performances majeures est l'Intelligence Artificielle. Nous devrions plutôt parler de multiples intelligences artificielles, autant d'algorithmes, de réseaux neuronaux et d'outils entraînés pour accélérer des tâches longues à réaliser dans le développement du produit. Certaines de ces tâches sont en amont de la simulation numérique pour générer des concepts, d'autres sont en substitution ou en complément de la simulation pour accélérer ou améliorer le résultat. Grâce à cette technologie, les boucles de conceptions pourraient être accélérées.

L'IA peut aussi aider sur de nombreuses autres tâches : la recherche d'informations (standard, pièces similaires, données qualité...), la génération de ligne de codes pour les systèmes embarqués, la définition et la réalisation des tests de validation, etc.

L'ensemble des apprentissages et des inférences pour utiliser ces IA nécessite encore d'importantes ressources calculs. Le maintien de notre compétitivité passe forcément par des infrastructures informatiques de haut niveau, agile pour allouer des ressources massives en fonction des besoins. La compétitivité des infrastructures de calculs est une condition nécessaire à la compétitivité de notre industrie.

En conclusion, la simulation numérique et, plus globalement, la transformation digitale de l'Ingénierie sont au cœur de l'accélération des développements véhicules et permettent de concevoir des produits innovants et performants au meilleur coût pour nos clients. ■

HPC ET IA AU SERVICE DE LA SUPÉRIORITÉ NAVALE

Cédric Leblond

Maître d'Ouvrage du Calcul Scientifique, Naval Group

Naval Group conçoit, réalise, intègre, maintient en service, modernise, démantèle et déconstruit des sous-marins et navires de surface, pour la Marine française et 50 marines alliées. Ce positionnement est unique puisque l'entreprise intervient sur l'ensemble du cycle de vie des navires armés, des pré-études au design jusqu'à la déconstruction en passant par l'intégration physique et fonctionnelle des systèmes.

L'entreprise est héritière de près de quatre cents ans d'innovations, contribuant à la supériorité technologique de ses clients et à l'amélioration de sa compétitivité. L'innovation est traduite aujourd'hui en six axes majeurs : Smart Naval Force, Invulnerable Ship, Smart Ship, Blue Ship, Smart Availability et Smart Industry.

La simulation numérique au sens large est aujourd'hui un outil majeur et transverse dans l'entreprise, présente sur la totalité du cycle de vie des navires. C'est également un véritable levier d'innovation dans chacun des six axes.

ÉVOLUTION DE L'USAGE DU HPC

Depuis plus de 30 ans, le calcul haute performance (HPC) est central en phase de conception détaillée des navires, notamment pour les simulations multi-échelles et multi-physiques. Son usage s'étend désormais à l'ensemble du cycle de développement. Dès la phase d'offre, il permet de vérifier les performances attendues ou comparer plusieurs variantes. En phase de conception amont, il facilite l'intégration rapide d'innovations et contribue à l'atteinte

du « bon du premier coup », facteur clé pour optimiser coûts, délais et qualité sur l'ensemble du programme.

Les besoins en conception détaillée se sont intensifiés : la simulation fine est désormais systématiquement utilisée pour vérifier la tenue des exigences à chaque itération, selon les principes de Vérification et Validation. Par ailleurs, les échanges entre concepteurs et calculateurs s'effectuent en quasi-temps réel, rendus possibles par la massification du HPC, garantissant réactivité, robustesse et fiabilité de la conception.

Le HPC intervient aussi en production. La simulation du soudage justifie les pratiques industrielles et optimise les séquences d'assemblage. La simulation des procédés de fabrication additive permet de définir les bons paramètres, réduisant essais, coûts et délais. Des analyses fluide-structure sont menées pour vérifier la tenue sismique des navires. En maintien en condition opérationnelle, les plateformes sont modernisées pour répondre à de nouvelles menaces et exigences capacitaires. L'intégration de ces innovations sous fortes contraintes de temps positionne le HPC comme un levier essentiel d'aide à la décision, garantissant à la fois performance technique et disponibilité opérationnelle.

ENJEUX À COURT ET MOYEN TERMES

Les enjeux du HPC dans le naval de défense sont multiples. La multiplication des tensions régionales impose d'accélérer les cadences de production, avec des phases d'étude plus courtes, en conservant un niveau équivalent

de fiabilité et de robustesse. Cela implique d'augmenter la puissance de calcul et d'assurer une flexibilité accrue via l'externalisation, tout en respectant des exigences strictes de cybersécurité, souveraineté technologique et conformité aux standards de défense. Le traitement de données sensibles impose des infrastructures maîtrisées, résilientes et souveraines, sans compromettre la scalabilité ni la réactivité opérationnelle.

L'évolution des menaces et la perspective de conflits de haute intensité redéfinissent les exigences capacitaires des bâtiments de combat. Les plateformes doivent être livrées prêtes à l'engagement, interopérables, adaptables et maîtrisées en coût. Une solution passe par le développement de jumeaux numériques à l'échelle du navire, rendu possible par le HPC et le machine learning. Ces jumeaux permettent de simuler, optimiser et valider virtuellement l'architecture et les systèmes du navire avant leur réalisation, réduisant les risques techniques et les délais de mise en service. En opération, ils offrent aux marines des capacités d'analyse en temps réel, de diagnostic et d'optimisation des

performances, renforçant ainsi l'avantage tactique et la résilience des forces navales.

L'intégration de l'IA dans les processus d'ingénierie, et sa convergence avec le HPC, devient également essentielle. L'IA transforme chaque étape : génération automatique des plans d'expérience, assistance à la mise en données, orchestration des calculs, accélération des solveurs, analyse des résultats, valorisation des données. Cette évolution nécessite une acculturation des experts HPC à l'IA et le recrutement de profils à double compétence. Dans la mesure où l'IA mobilise une part croissante des budgets, redéfinit les standards techniques (suprématie du GPU, retour à la simple précision) et attire les profils les plus recherchés, une marginalisation du secteur HPC traditionnel est possible. L'intégration maîtrisée de l'IA et sa fusion avec le HPC sont donc un enjeu primordial de préservation des savoir-faire, d'attractivité et de compétitivité. Plus généralement, la maîtrise conjointe du HPC et de l'IA est un levier essentiel pour assurer la souveraineté technologique et la supériorité opérationnelle navale. ■



Illustration du navire numérique, véritable clone numérique d'un bâtiment existant. Copyright: Naval Group.

ENERGIE NUCLÉAIRE : L'INDISPENSABLE LIEN AU NUMÉRIQUE ET AU CALCUL INTENSIF

Patrick Blanc-Tranchant

Directeur adjoint des programmes énergies, CEA

UNE NOUVELLE DYNAMIQUE À L'ŒUVRE SUR LE NUCLÉAIRE...

Face à l'urgence climatique le nucléaire s'impose comme une composante essentielle de l'impérative transition énergétique des sociétés, grâce à sa capacité à produire massivement une électricité décarbonée et pilotable et aux perspectives qu'il ouvre, au-delà de l'électricité, pour décarboner la chaleur industrielle, produire de l'hydrogène bas carbone ou encore des carburants de synthèse.

En France, la reconnaissance de son rôle stratégique pour l'atteinte de la neutralité carbone impulse, depuis 2020, une nouvelle dynamique politique et industrielle qui vise la prolongation de l'exploitation du parc historique, le déploiement de réacteurs de puissance destinés à prendre le relais le moment venu, mais aussi le développement de réacteurs très innovants qui pourraient bénéficier d'une nouvelle approche en matière de construction, étant réalisés en série et pour partie directement en usine (réacteurs de type « SMR », pour Small Modular Reactors).

... NÉCESSITANT UN APPUI RENFORCÉ DU NUMÉRIQUE ET DE LA SIMULATION AVANCÉE

La modélisation puis la simulation ont accompagné le développement du nucléaire depuis son émergence mais la nouvelle dynamique qu'il connaît va appeler dans l'avenir à un recours absolument sans précédent à ces outils et techniques, et plus largement au numérique.

L'exploitation des données, le développement de jumeaux numériques, qui s'appuient sur des techniques et des approches entrant dans le champ de l'Intelligence Artificielle (IA), vont permettre d'optimiser la maintenance prédictive, d'améliorer le pilotage des réacteurs, d'anticiper les aléas, ... En un mot, le numérique va (continuer à) transformer en profondeur les pratiques industrielles.

De nombreux nouveaux champs d'application de la modélisation/simulation vont par ailleurs s'ouvrir :

- La prolongation des réacteurs existants nécessitera des études approfondies, en complément de celles conduites jusqu'ici, de façon à pouvoir justifier le comportement des matériaux et du combustible ou la tenue mécanique des composants par exemple,
- Les nouveaux réacteurs vont aussi amener des besoins extrêmement importants, pour leur conception, leur dimensionnement ou en appui des démonstrations de sûreté qui seront nécessaires avant qu'ils ne voient le jour. Les nouveaux combustibles qu'ils utiliseront, en particulier, vont amener de très forts besoins d'études,
- Enfin, l'intégration du nucléaire dans un mix énergétique plus diversifié que par le passé va également amener à des besoins nouveaux : besoin de réacteurs plus manœuvrant, ce qui aura des répercussions notamment sur leur combustible, et pour qu'ils puissent contribuer à la production de chaleur ou d'hydrogène par exemple cela va, de même, nécessiter énormément d'études nouvelles.

Toutes ces évolutions, majeures, vont venir renforcer le besoin pour une simulation « haute-fidélité » des réacteurs, au plus près de la physique sous-jacente.

CONCRÈTEMENT : LA SIMULATION HAUTE-FIDÉLITÉ...

Ce besoin d'une simulation avancée, plus prédictive, nécessitera de renforcer les approches dites « multi-échelles » et « multiphysiques » : mieux comprendre et modéliser les phénomènes physiques élémentaires, les combiner par des démarches de « remontée d'échelles » et mieux faire interagir des disciplines physiques qui aujourd'hui sont encore le plus souvent traitées séparément, à partir de modèles propres à chaque domaine.

Les progrès attendus doivent porter sur toutes les disciplines physiques clés mises en jeu dans le cadre du fonctionnement des réacteurs, la neutronique, associée à la physique nucléaire, la radioprotection, la mécanique des fluides, la thermomécanique des structures, la science des matériaux (simulation du combustible et des autres constituants des installations), la chimie (simulation de la corrosion notamment), ainsi que sur leur couplage, compte tenu de l'interdépendance de fait des phénomènes physiques à modéliser.

Ceci se complète de la volonté de progresser sur l'évaluation des incertitudes associées aux simulations, dans un objectif d'optimisation, par réduction des marges que ces incertitudes imposent de prendre aujourd'hui, ce que permettront de nouveaux moyens et de nouvelles approches qui entrent là aussi dans le champ de l'IA (utilisation de réseaux de neurones, de méta-modèles, etc.).

Les travaux correspondants sont déjà en cours de déploiement. Dans les différentes disciplines physiques des simulations progressivement plus fines sont mises en œuvre : recours plus systématique aux techniques Monte-Carlo pour la théorie du transport en neutronique, simulations fines de type CFD en thermohydraulique par exemple, loin des approximations historiquement nécessaires. Et le couplage entre ces différentes disciplines progresse de même à grands pas. Mais l'ensemble

va devoir être considérablement amplifié pour passer de calculs encore relativement limités en taille à la possibilité de traiter l'échelle des réacteurs complets.

... ET LA NÉCESSITÉ DU RECOURS AU CALCUL HPC

Continuer à progresser ainsi, d'abord pour réaliser des calculs de référence qui serviront d'appui aux calculs industriels puis, le moment venu, pour contribuer directement à ces calculs industriels, reposera pour beaucoup sur les puissances de calcul accessibles.

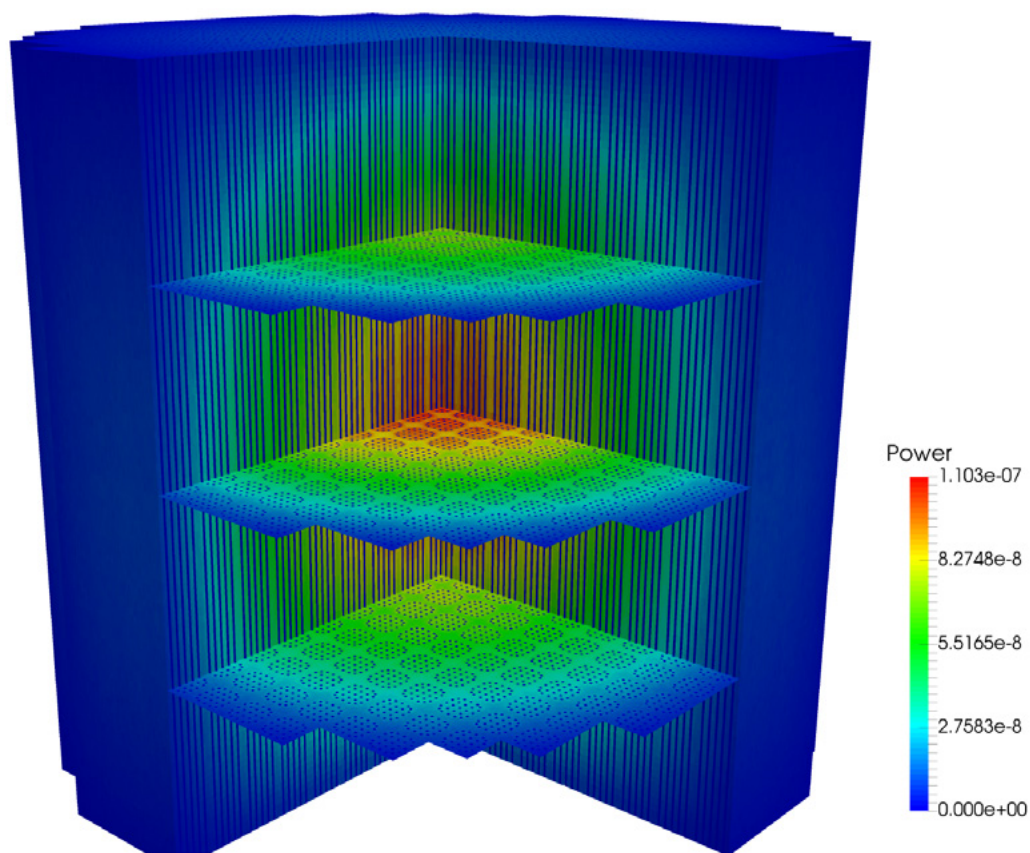
Pour en prendre la mesure, nous avons évalué les besoins pour simuler un transitoire accidentel de réacteur REP (réacteur actuel) avec une finesse de modélisation en rupture par rapport à ce qui se fait classiquement aujourd'hui pour ce type d'études, et en couplant intimement les trois disciplines majeures que sont la neutronique, la thermohydraulique et la physique matériaux/combustibles, là aussi en rupture par rapport aux études courantes qui ne traitent pas encore simultanément ces trois physiques différentes.

L'évaluation a montré que pour atteindre le degré de finesse visé, qui constituerait un progrès majeur, de l'ordre de 500 millions d'heures de calcul (CPU) seraient nécessaires ; à comparer aux calculs les plus fins menés aujourd'hui, qui mobilisent plutôt de l'ordre de quelques millions d'heures de calcul seulement. Deux ordres de grandeur donc.

ET PLUS LARGEMENT ?

Pour élargir la vision soulignons que ces apports renforcés attendus de la simulation et du calcul haute performance pour les systèmes nucléaires le seront de même dans l'avenir pour les autres moyens de production ou de stockage d'énergie (piles à combustible ou batteries par exemple) et plus largement pour accompagner les transformations de notre système énergétique global aujourd'hui à l'œuvre, pour son dimensionnement, son pilotage ou le suivi de ses performances.

Demain plus encore qu'aujourd'hui la simulation et le calcul haute performance pour l'énergie seront des indispensables. ■



Visualisation issue d'un calcul détaillé d'une nappe de puissance d'un cœur de réacteur complet à l'aide du logiciel PATMOS, pré-figurateur des futures simulations Monte-Carlo en physique des réacteurs du CEA/DES portées par le code TRIPOLI5, aujourd'hui en développement (benchmark dit Hoogenboom, mis en place dans le cadre de l'OCDE/AEN - <https://inis.iaea.org/records/3jkzd-j6f83>). Le calcul correspondant était un très gros calcul sur les moyens HPC CEA, en rupture par rapport aux simulations antérieures, pour lesquelles la finesse atteinte ici n'était pas accessible à l'échelle d'un cœur complet de réacteur.

LES GRANDS ENJEUX DE LA TRANSITION VERS LA NEUTRALITÉ CARBONE ET RÔLE DE LA TRANSFORMATION NUMÉRIQUE

Alain Martin

Responsable du Système d'Information Scientifique, EDF R&D

Stéphane Tanguy

Directeur des Systèmes et Technologies de l'Information, EDF R&D

Contexte : le calcul scientifique et la simulation sont des outils essentiels pour assurer le bon fonctionnement du système électrique d'aujourd'hui et pour construire son futur.

Depuis plus de 30 ans, EDF s'appuie sur le calcul scientifique et la simulation pour garantir la sûreté et la performance de ses installations, assurer l'équilibre offre demande avec des prévisions très fiables de consommation et répondre aux multiples défis de la transition énergétique.

Si EDF exploite ses propres supercalculateurs (HPC) classés parmi les 100 plus puissants au monde, la R&D du Groupe a également développé son propre portefeuille de codes de simulation modélisant les différentes physiques et garantissant la représentativité de nos simulations. Certains de ces grands codes sont disponibles en open source et utilisés par une communauté scientifique qui dépasse largement le cadre national.

Nous avons introduit plus récemment des algorithmes d'Intelligence artificielle dans nos codes de façon à produire plus rapidement et simplement des résultats sans dégradation notable de la précision (par exemple pour développer un métamodèle qui est utilisé pour fournir une carte de chaleur qui servira à prioriser des réparations de soudures).

Depuis quelques années, nous nous intéressons de très près à l'algorithmie quantique qui jouera un rôle prépondérant lorsque les ordinateurs quantiques à

l'échelle seront disponibles et nous permettront de traiter des problèmes là où la simulation classique ne fournit que des approximations.

SUPERCALCULATEUR : OUTIL CLÉ ET PILIER DE LA R&D POUR LA MODÉLISATION ÉNERGETIQUE

Avec une puissance de calcul dépassant 11 pétaflops (soit 11 millions de milliards d'opérations par seconde), les supercalculateurs d'EDF constituent un levier essentiel de la R&D et sont au cœur des efforts d'innovation d'EDF. Grâce à eux, l'entreprise modélise avec une précision accrue des phénomènes physiques complexes tout en s'appuyant sur une base expérimentale dédiée pour valider et qualifier ses propres codes de calcul, garantissant ainsi la fiabilité et la pertinence de ses simulations.

A titre d'exemple, ces ressources sont mobilisées pour :

- La simulation des phénomènes physiques qui régissent la performance et la sûreté de nos moyens de production (Nucléaire, thermique, hydraulique ou renouvelables) grâce à des codes de physique des champs comme `code_saturne`, `neptune_cfd`, `code_aster` et `telemac` ;
- La prévision de la consommation et l'optimisation des moyens de production pour garantir l'équilibre offre demande à tout instant ;
- L'anticipation des impacts du changement climatique afin de garantir la continuité d'exploitation de nos moyens actuels pour les décennies à venir.

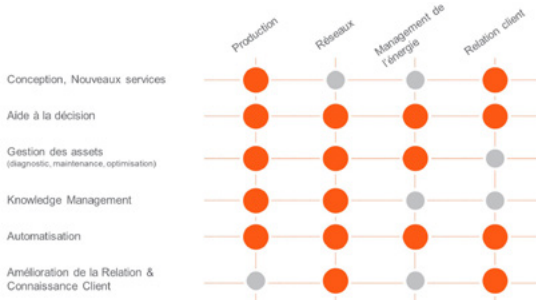
LES GRANDS ENJEUX DE LA TRANSITION VERS LA NEUTRALITÉ CARBONE ET RÔLE DE LA TRANSFORMATION NUMÉRIQUE

L'IA DÉJÀ LARGEMENT DÉPLOYÉE DANS LE SYSTÈME ÉLECTRIQUE

L'IA est devenue un levier stratégique incontournable pour EDF. Initialement maîtrisée par la R&D, elle s'est progressivement déployée à travers le Groupe, notamment grâce aux Usines Data Analytics, qui accompagnent les métiers dans le développement de cas d'usage à forte valeur ajoutée pour les besoins suivants :

Avec l'essor des IA génératives, EDF explore de nouvelles opportunités tout en veillant à une intégration responsable et sobre de ces technologies, afin de garantir leur transparence, leur sécurité et leur faible impact énergétique.

L'INFORMATIQUE QUANTIQUE : UNE RUPTURE



TECHNOLOGIQUE PROMETTEUSE POUR EDF

Le calcul quantique représente l'avenir du calcul scientifique et la promesse d'avancées scientifiques majeures inaccessibles par le calcul classique. Il représente également une rupture complète d'un point de vue algorithmique il est donc primordial d'anticiper le portage de nos codes de simulation vers cette nouvelle technologie.

EDF explore actuellement plusieurs catégories de cas d'usage prometteurs :

- L'optimisation : amélioration du pilotage du smart charging pour une gestion plus efficace des réseaux électriques ;
- Les matériaux : modélisation à l'échelle moléculaire afin de mieux comprendre le vieillissement des matériaux sous irradiation, et d'optimiser la durée de vie des batteries et des cellules photovoltaïques ;



HPC CRONOS en Data Center EDF

- La simulation de la physique par les équations aux dérivées partielles (EDP) : des gains significatifs en performance et l'accès à des résultats de simulation inatteignables avec les approches classiques.

En collaboration avec des partenaires académiques et industriels, EDF anticipe l'intégration du quantique dans ses méthodes de simulation comme dans ses infrastructures pour rester à la pointe de l'innovation. Il s'agit d'un enjeu stratégique, tant pour renforcer sa compétitivité que pour répondre aux exigences d'un secteur énergétique en pleine mutation.

L'IA ET LE CALCUL QUANTIQUE SONT DES LEVIERS ESSENTIELS POUR RELEVER LES DÉFIS DE NOTRE FUTUR SYSTÈME ÉLECTRIQUE.

Si comme on l'a vu la simulation numérique a représenté un outil essentiel pour assurer performance et sûreté de notre système électrique, les nouveaux défis associés à la transition énergétique, au changement climatique et à la décarbonation de notre économie vont nécessiter plus que des PFlops supplémentaires. Il s'agit désormais de produire de nouvelles modélisations plus complexes utilisant des données jusque-là inexploitées qui ne se feront pas sans les apports de l'Intelligence artificielle et du calcul quantique. ■

IA HYBRIDE AU SERVICE DES VILLES ET TERRITOIRES CONNECTÉS, SOUTENABLES ET RÉSILIENTS

Francisco Chinesta

Professeur des universités, Arts et
Métiers Sciences et Technologies

Dominique Baillargeat

Directeur exécutif scientifique
CNRS@ CREATE (Singapour)

Victor Champaney

Co-fondateur,
Duoverse

La principale particularité d'une ville ou d'un territoire, en matière de prédictibilité, est leur complexité extrême.

L'existence de nombreux systèmes interconnectés en interaction, de grande taille spatiale et temporelle, présentant différentes granularités (systèmes de systèmes) et traversés par des incertitudes et des imprécisions qui s'y propagent, rend la modélisation particulièrement complexe. Si l'on y ajoute le principal protagoniste, l'humain, dont le comportement échappe aux modèles déterministes, l'espoir d'atteindre le niveau de diagnostic, de pronostic et d'aide à la décision souhaité s'en trouve considérablement limité.

MODÉLISER POUR MIEUX PRÉDIRE

Les modèles issus des sciences de l'ingénieur, conçus pour prédire le comportement d'un composant ou d'un système multi-composant, ont connu une progression colossale, tant en précision qu'en efficacité, cette dernière étant tributaire des avancées tout aussi

considérables en capacité de calcul, portées par la démocratisation du calcul haute performance (HPC).

Les techniques de discrétisation ont également joué un rôle clé, permettant la résolution efficace de modèles de plus en plus grands, à l'échelle spatiale comme temporelle. À la fin du siècle dernier, les techniques de réduction de modèles ont permis de résoudre, souvent en temps réel, des problèmes multi-physiques, multi-échelles, non linéaires, couplés et de grande taille. Aujourd'hui, vitesse et précision ne sont plus incompatibles !

En ce début de XXI^e siècle, la réalité physique s'est rapidement orientée vers une numérisation, avec pour objectif une prédiction à la fois rapide et précise, afin d'anticiper et de prendre des décisions optimales, aussi bien lors de la conception que durant l'exploitation des systèmes d'ingénierie.

C'est dans ce contexte, que le jumeau numérique qui est une réplique virtuelle de la réalité observée, a vu le jour. Parfois utilisé dès la phase de conception, il



Fig. 1: Réalité augmentée sur maquette (en collaboration avec Duoverse et IMMERSION)

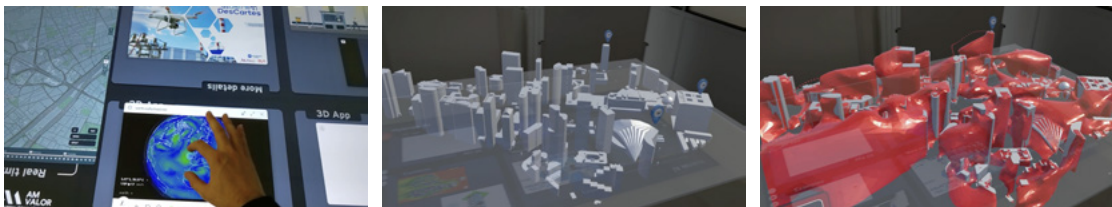


Fig. 2: Réalité virtuelle sur table d'ingénierie collaborative (en collaboration avec Duoverse et IMMERSION)

s'appuie sur les connaissances accumulées, les données historiques et l'ensemble des acquis des sciences de l'ingénieur. D'autres fois, il vise à optimiser l'exploitation des systèmes déjà fabriqués et en service, tout au long de leur cycle de vie. Ainsi se trouvent renforcées les deux missions fondamentales de l'ingénieur : concevoir et opérer.

Après quatre grandes révolutions — la vapeur, l'électricité, l'électronique (automatisation), puis la donnée avec l'IA —, la cinquième intégrera l'humain, longtemps relégué au rôle de spectateur plus que d'acteur, dans les villes et les territoires. Il ne s'agit plus de tenter d'automatiser le fonctionnement des villes, une tentative déjà faite et vite abandonnée face à ses échecs, mais de développer des technologies à faible empreinte environnementale et énergétique, transparentes pour l'utilisateur, et au service du citoyen : pour son confort, son plaisir, son bien-être, sa sécurité et son efficacité.

DU DIAGNOSTIC À L'ACTION : VERS UN JUMEAU NUMÉRIQUE HYBRIDE

Une telle technologie aura trois niveaux d'action : (I) une action déportée, déclenchée à partir d'un diagnostic et d'un pronostic ; (II) une aide à la décision pour les autorités, basée sur l'analyse en temps réel d'innombrables scénarios ; et (III) une information accessible au citoyen, lui permettant d'adapter ses choix, tant dans la vie quotidienne qu'en situation d'urgence.

Cependant, si le diagnostic peut être réalisé par une simple analyse des données à l'aide des techniques d'IA, le pronostic, quant à lui, exige une projection vers le futur, une capacité propre aux modèles !

Cela étant dit, les modélisations actuelles appliquées aux systèmes complexes révèlent très souvent un écart notable entre les prédictions et les observations. La

réalité semble contenir des éléments que les modèles ne parviennent pas à capturer, en raison d'imprécisions ou d'incertitudes, parfois d'ordre épistémique.

L'approche « tout données » n'est pas non plus viable, car elle nécessiterait une quantité massive de données pour décrire avec une fine granularité les villes ou les territoires. Imaginez devoir installer des millions de capteurs dans chaque bâtiment, chaque rue, chaque infrastructure...

Les deux paradigmes montrent leurs limites, ouvrant la voie à un paradigme de modélisation hybride. Gardons les capteurs, en nombre réduits mais placés dans des positions optimales. Pour cela, les modèles issus de la physique et des connaissances sont d'une aide précieuse ! Complétons les données avec les modèles des sciences de l'ingénieur offrant une estimation, certes pas parfaite, mais pas aberrante non plus. Les données, exploitées par l'IA, viendront ainsi enrichir ces modèles de base, pour former un modèle dit hybride. Une belle alliance entre précision, rapidité, frugalité (peu de données), et explicabilité (essentielle pour certifier) !

Ce jumeau, dit hybride, permet de concevoir fidèlement, mais aussi d'exploiter des systèmes qui échappent aux anciens paradigmes de modélisation.

DESCARTES, LABORATOIRE DE LA VILLE AUGMENTÉE

Dans le cadre des villes, le projet DesCartes (<https://descartes.cnrsatcreate.cnrs.fr/>) porté par la filiale CNRS@CREATE du CNRS à Singapour, au sein du campus d'excellence CREATE, et financé par la NRF (National Research Foundation) propose une méthodologie de modélisation hybride pour plusieurs thématiques clés : (I) l'environnement ; (II) la surveillance des structures ; (III) les réseaux intelligents ; (IV) la mobilité, notamment autour des drones ; et (V) la gestion de crises ciblant notamment la ville, le port et

l'aéroport. Cette méthodologie est conçue pour être déployable sur d'autres villes et territoires, au-delà du contexte singapourien.

Ainsi, la modélisation paramétrique du vent en milieu urbain, avec une résolution métrique, à partir des données de vent en altitude (direction et intensité) et de quelques stations météorologiques, permet d'estimer plusieurs phénomènes clés : les trajectoires optimales des drones, la distribution de température avec l'identification des îlots de chaleur, la qualité de l'air à partir des données de circulation en temps réel, ou encore la dispersion de polluants issus du port (par exemple des navires) ou d'émissions accidentelles provenant de sites industriels.

L'impact de ces phénomènes peut ainsi être estimé en temps réel avec une grande précision, permettant d'agir de manière adéquate.

Cette modélisation hybride nécessite la convergence de plusieurs technologies : (I) des techniques de réduction de modèles paramétriques, appliquées à des phénomènes physiques tels que la mécanique des structures, les écoulements, les transferts thermiques ou encore l'électromagnétisme (par exemple pour la couverture satellite ou 5G); (II) des méthodes d'apprentissage machine, utilisées pour enrichir les modèles fondés sur la physique; (III) des techniques d'assimilation de données, permettant de calibrer les modèles à partir d'observations réelles.

Le tout s'appuie sur de puissantes infrastructures de calcul haute performance (HPC), et plus récemment, sur les capacités du calcul quantique (notamment le quantum annealing) pour l'extraction de modèles réduits, la résolution de problèmes inverses (comme l'identification de sources ou la recherche de root-cause), l'affectation multidimensionnelle ou encore la coordination d'agents collaboratifs. Les résultats obtenus dans ces domaines sont à ce jour très prometteurs. ■

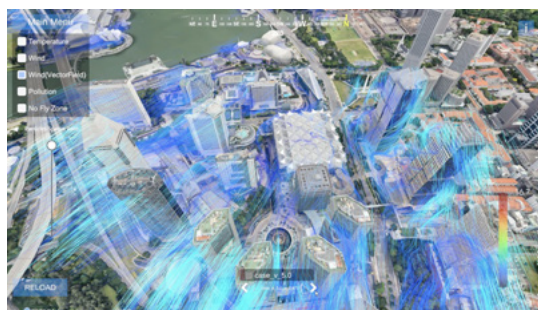


Fig. 3: Réalité virtuelle immersive : Modèle (haut), température et pollution (centre et bas) - En collaboration avec Duoverse et IMMERSION

ACCÉLÉRER LA DÉCOUVERTE DE NOUVEAUX MATÉRIAUX : LA CONVERGENCE DU HPC ET DE L'IA

Jean-Yves P. Delannoy

Scientific Director Digital R&D, Arkema

La découverte et le développement de matériaux avancés ont traditionnellement été une entreprise longue et gourmande en ressources, tributaire d'expérimentations par essais erreurs. Cependant, ce paradigme évolue rapidement en raison de la convergence du Calcul Haute Performance (HPC) et de l'Intelligence Artificielle (IA). Ces technologies transforment une innovation matériaux vitale pour relever les défis mondiaux dans les domaines de l'énergie, de la durabilité et de la fabrication.

Historiquement, il pouvait falloir des décennies pour qu'un matériau passe du concept initial à son déploiement sur le marché. Aujourd'hui, l'intégration du HPC et de l'IA offre une voie pour accélérer considérablement ce processus, réduire les coûts de développement et explorer des domaines de l'espace chimique auparavant inaccessibles.

L'ÉVOLUTION DE LA DÉCOUVERTE DE MATÉRIAUX

La découverte traditionnelle de matériaux reposait sur des flux de travaux manuels, séquentiels et à fort besoin de main-d'œuvre. Les chercheurs synthétisent et testent plusieurs candidats individuellement, puis itèrent en fonction des résultats obtenus. Bien que fondamentale, cette méthode est limitée par l'intuition humaine, les ressources et le vaste espace des combinaisons de matériaux accessible. Désormais l'automatisation, l'expérimentation en parallèle, la modélisation et la simulation pilotées par l'IA permettent de réduire cette complexité¹. Malgré les défis

persistants – comme la garantie de la reproductibilité – ces innovations transforment la science des matériaux. Le calcul haute performance, la modélisation multi-échelle et l'IA sont au cœur de ce changement. Les outils de calcul avancés permettent le criblage à haut débit de vastes espaces chimiques, améliorant ainsi la conception prédictive des matériaux.

CALCUL HAUTE PERFORMANCE ET SIMULATION

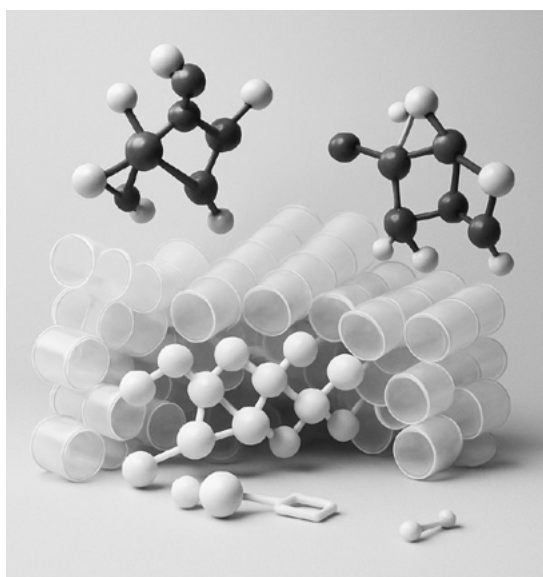
Le HPC permet aux chercheurs de simuler les propriétés des matériaux avec une précision croissante. Les méthodes *ab initio*, basées sur la mécanique quantique, modélisent les interactions atomiques et les propriétés des matériaux avant leur synthèse, réduisant ainsi les coûts et augmentant l'efficacité de la recherche. Les simulations de dynamique moléculaire ajoutent un aperçu dynamique du comportement des matériaux. Ces outils sont particulièrement efficaces dans les domaines du stockage de l'énergie et de la catalyse. Avec l'introduction du cloud, l'accès au supercalculateur a été élargi et démocratisé, accélérant ainsi l'innovation. En science des polymères, les simulations guident désormais la conception macromoléculaire et réduisent le travail en laboratoire. Concilier le besoin de précision avec l'efficacité et la rapidité des calculs reste néanmoins un défi. Les modèles à gros grains aident à étudier les systèmes complexes sur des échelles de temps plus longues, tandis que l'IA repousse les limites de ce qui est faisable sur ordinateur et interprétable en science des matériaux².

INTELLIGENCE ARTIFICIELLE DANS LA DÉCOUVERTE DE MATÉRIAUX

L'IA transforme la découverte de matériaux en générant de nouvelles hypothèses, en concevant de nouveaux matériaux et en accélérant la prise de décision³. Avec l'essor de l'IA générative et des grands modèles de langage (LLM), des outils tels que MatterGen de Microsoft et GNoME (Graph Networks for Materials Exploration) de Google peuvent proposer de nouvelles structures moléculaires optimisées pour des propriétés spécifiques. Les modèles d'apprentissage automatique (ML) classiques contribuent également en permettant une prédiction rapide des propriétés, comme le démontre le développement de polymères durables.

Les modèles de ML – en particulier les architectures d'apprentissage profond – excellent dans la compréhension de relations complexes entre la structure et les propriétés. Ces modèles réduisent considérablement l'espace de recherche expérimental, diminuent le temps nécessaire à la découverte et améliorent la reproductibilité.

Il est important de noter que l'IA atténue le compromis traditionnel entre précision des simulations et vitesse de calcul. Ce changement de paradigme positionne l'IA comme un catalyseur central pour résoudre les défis critiques dans les domaines de l'énergie, de la durabilité et de la fabrication avancée.



HPC ET IA : ACCELERATEURS DE NOUVEAUX MATERIAUX

La convergence du Calcul Haute Performance et de l'Intelligence Artificielle marque une ère de transformation dans la découverte de matériaux innovants. Ensemble, ces technologies offrent des capacités inégalées pour concevoir et développer à un rythme accéléré, des matériaux aux propriétés sur mesure.

En exploitant la puissance de calcul du HPC et de l'IA, les chercheurs peuvent désormais naviguer dans de vastes paysages chimiques avec une efficacité auparavant inimaginable. À mesure que ces technologies mûrissent et deviennent de plus en plus accessibles, les organisations qui les adoptent se positionneront à l'avant-garde de l'innovation des matériaux des prochaines générations.

L'avenir de la science des matériaux ne réside pas dans le choix entre le HPC ou l'IA, mais dans la synergie de leurs forces pour accélérer le passage du concept à la commercialisation de solutions innovantes. ■

Références

1. Edward O Pyzer-Knapp, Jed W Pitera, Peter WJ Staar, Seiji Takeda, Teodoro Laino, Daniel P Sanders, James Sexton, John R Smith, and Alessandro Curioni. Accelerating materials discovery using artificial intelligence, high performance computing and robotics. *npj Computational Materials*, 8(1):84, 2022.
2. Thomas E Gartner III and Arthi Jayaraman. Modeling and simulations of polymers: a roadmap. *Macromolecules*, 52(3):755–786, 2019.
3. Yuanqi Du, Arian R Jamasb, Jeff Guo, Tianfan Fu, Charles Harris, Yingheng Wang, Chenru Duan, Pietro Liò, Philippe Schwaller, and Tom L Blundell. Machine learning-aided generative molecular design. *Nature Machine Intelligence*, 6(6):589–604, 2024.

Fig. 1: Artistic Illustration of cellulose acetate plasticization - Generated using ChatGPT 4o

IA, CALCUL INTENSIF ET QUANTIQUE AU SERVICE DES FORCES DE DEMAIN

Emmanuel Chiva

Délégué Général pour l'Armement, Ministère des Armées

Michaël Krajecki

Directeur du pôle Recherche, Agence Ministérielle pour l'IA de Défense

20 ANS DÉJÀ!

Teratec est devenu un acteur français et européen incontournable de la simulation numérique, du big data et de l'IA! Alors oui, les mots-clefs de la décennie sont IA, apprentissage profond, LLM, jumeaux numériques... Toutes ces avancées ont fait irruption dans notre quotidien et sont à l'origine de nombreux défis techniques, organisationnels ou éthiques. L'exercice de prospective est toujours difficile et il est bien souvent plus simple de regarder dans le rétroviseur plutôt que d'anticiper le prochain virage en contrebas de la route. La dernière décennie a été très riche pour Teratec, nous avons été parfois spectateurs, souvent acteurs de ces transformations. Le Ministère des Armées s'est fortement mobilisé à vos côtés: par l'intermédiaire de l'Agence de l'Innovation de Défense, il soutient de nombreux projets portés par des chercheurs de nos laboratoires, par des start-up ou encore de grands industriels qui participent tous au rayonnement de Teratec.

Technologie duale, l'IA - après avoir traversé plusieurs hivers - a connu sa révolution grâce à la conjonction de trois facteurs. Tout d'abord, l'information est largement disponible et nativement produite sous forme numérique, permettant la constitution de corpus de données gigantesques. Mais surtout, les puissances de calcul largement accessibles permettent une mise en œuvre des mécanismes d'apprentissage sur lesquels repose l'IA développée sur la dernière décennie. Enfin, les avancées algorithmiques récentes, parmi

lesquelles nous pouvons retenir l'apprentissage profond ou encore les réseaux adverses et les mécanismes d'attention, ont abouti au développement de nouvelles applications intelligentes et souvent embarquées. Au sein du Ministère des Armées, l'Agence Ministérielle pour l'IA de Défense (AMIAD) produit et développe les technologies d'IA au service des forces. La Direction Générale de l'Armement (DGA) porte l'implémentation de ces IA au cœur des systèmes d'armes.

Alors qu'AlexNet avait déjà révolutionné le monde de l'image en 2012, l'avènement des GPU se confirmait de TOP500 en TOP500. En novembre 2013, l'université de Reims - membre de Teratec - classe son ROMEO dans le TOP500 et dans le GREEN500: pour la première fois, un supercalculateur accéléré français est disponible pour la recherche académique. En 2017, alors que Stanford publie la première édition du AI Index Report, le député Cédric Villani, médaillé Fields en 2010, a été chargé de conduire une mission sur la mise en œuvre d'une stratégie française et européenne en IA, qui se poursuit aujourd'hui encore dans le cadre de France 2030. À l'été 2022, des chercheurs issus des organismes de recherche et du monde des start-up bénéficient du supercalculateur Jean Zay pour développer le modèle GPT3 BLOOM. Fin 2022, ChatGPT réunit 1 million d'utilisateurs en 5 jours et la dynamique s'amplifie. En France, Mistral - société créée en avril 2023 - propose des modèles dits open weight.

La convergence entre données, algorithmes et puissance de calcul a révolutionné notre monde en quelques années d'une façon vertigineuse !

À cette révolution s'en ajoute une autre, la révolution quantique : capteurs, cryptographie et calcul... Du point de vue de la pertinence opérationnelle, la DGA concentre ses efforts sur les capteurs quantiques (horloges atomiques, gravimètres à atomes froids). Cependant, en ce qui concerne le calcul, le Ministère des Armées considère que la recherche civile sera en mesure d'accélérer et de répondre à ses besoins. En 2018 déjà, ATOS proposait sa solution de simulation, la QLM (Quantum Learning Machine) qui permettait d'anticiper l'arrivée des premiers accélérateurs de calcul quantique.

Depuis 2022, le projet HQI (Hybrid HPC Quantum Initiative) propose une plateforme de calcul hybride couplant plusieurs technologies quantiques au supercalculateur Joliot Curie de GENCI au TGCC (CEA). Ce projet vise à favoriser l'émergence de nouveaux cas d'usage qui pourront tirer parti des calculateurs quantiques. Cette deuxième révolution devrait très certainement voir de nouvelles applications déboucher dans les années à venir.

Les acteurs de Teratec, académiques et industriels, sont au cœur de ces révolutions et bénéficient du soutien de France 2030 afin de faire rayonner la science développée en France, en Europe et partout à travers le monde.

COMMENT LE MINISTÈRE DES ARMÉES S'EMPRE-IL DE CES SUJETS ?

Les technologies d'IA joueront un rôle de premier plan dans la supériorité opérationnelle future des forces armées et pour autant, le monde de la Défense comporte des particularités qui justifient notre investissement constant dans le développement d'une IA d'emploi au service de nos armées et de la transformation du Ministère des Armées.

Outre les considérations classiques (éthique, cadre d'emploi, doctrine), le milieu militaire présente des singularités. La rareté des données : les données sont plus rares dans le domaine de la Défense, comportent une sensibilité particulière et reposent souvent sur des capteurs plus spécifiques au milieu militaire, comme des radars ou des sonars. Au-delà du problème de souveraineté posé par les clouds publics, ces derniers supposent une hyperconnectivité des systèmes, peu réaliste dans la Défense.

En 2025, le Ministère a mis en exploitation le plus puissant supercalculateur classifié dédié à l'IA en Europe. Ce dernier situé au Mont-Valérien à Suresnes, permet à la France de traiter souverainement des données confidentielles, pour le besoin des armées ainsi qu'au service des entreprises de défense. Il n'est pas connecté à internet et sa maintenance est réalisée par des citoyens français habilités au secret de la défense nationale.

De même, les applications du quantique pour la Défense sont nombreuses : centrales inertielles, nouvelles approches de chiffrement et applications du calcul... Ainsi, le programme PROQCIMA, financé par France 2030 et porté par la DGA, vise à développer en France deux prototypes d'ordinateurs quantiques universels tolérants aux fautes, chacun doté de 128 qubits logiques, et prêts à passer à l'échelle industrielle. Six start-up françaises sont ainsi dans la course au développement d'un tel système comme Pasqal, Alice et Bob ou encore Quandela...

ET DEMAIN ?

Pour le Ministère des Armées, renforcer ses compétences dans le domaine du quantique et de l'IA reste une priorité forte : dès 2027, en termes de recherche amont, le ministère va regrouper au sein de l'institut de recherche de défense, nouveau bâtiment totem de l'école polytechnique, les activités liées à la recherche en IA et en quantique pour la défense ainsi que le CIEDS (Centre Interdisciplinaire d'Études pour la Défense et la Sécurité).

IA, CALCUL INTENSIF ET QUANTIQUE AU SERVICE DES FORCES DE DEMAIN

Demain, les robots pourront sortir de nos usines pour interagir dans un environnement ouvert en lien direct avec le monde réel. Le robot deviendra un nouvel effecteur pour l'IA dans notre monde. En mars 2025, le Ministre des Armées a annoncé le projet Pendragon, qui vise à développer l'unité robotisée de combat de demain. Confiée à l'AMIAD et au Commandement du Combat futur, cette unité expérimentale sera capable de mener des actions tactiques de combat autonomes pour remplir des missions variées, avec pour finalité une adoption rapide par les forces et une industrialisation, et ce, dès 2027.

N'oublions pas que l'IA continue à se développer et a besoin de ressources de calcul exceptionnelles pour être toujours plus utile et précise. L'IA, ce sont des algorithmes et des données. Le Ministère des Armées sera à vos côtés pour porter l'innovation numérique française au meilleur niveau international. AID, AMIAD et DGA se mobiliseront pour que nos Armées disposent des meilleures technologies numériques pour assurer leurs missions. De Jean Zay, aux Giga factories européennes, les acteurs de Tera-tec ont toute leur place dans ces nouveaux défis! ■



Illustration du projet PENDRAGON – crédit : Ministère des Armées

LES TECHNOLOGIES DE CALCUL HAUTE PERFORMANCE POUR LE SECTEUR ASSURANCE-FINANCE

Christophe Michel

Responsable Validation et Risques de Modèle, Crédit Agricole CIB

LE LIEN FONDAMENTAL ENTRE HPC ET GESTION DES RISQUES FINANCIERS

Le calcul haute performance constitue un pilier stratégique pour l'industrie financière. La nature même des métiers bancaires et assurantiels repose sur une gestion sophistiquée des risques multidimensionnels : risques de dommage, risques de marché affectant la valorisation des éléments de bilan, risques de contrepartie, risques opérationnels, et risques de liquidité.

L'évaluation et la maîtrise de ces risques s'appuient sur des modèles mathématiques et actuariels dont l'implémentation nécessite d'importantes ressources informatiques, tant pour le traitement des données que pour l'exécution des calculs complexes.

Le cadre réglementaire post-crise 2008 a significativement intensifié ces besoins en calcul. L'introduction de directives comme Solvency 2 pour les assureurs et la Revue Fondamentale du Trading Book pour les activités de marché a imposé de nouveaux indicateurs de risque particulièrement exigeants en termes de puissance de calcul.

Un exemple emblématique est celui des activités de marché. La gestion du risque de marché s'articule principalement autour des produits dérivés. Leur valorisation et leur couverture reposent sur des méthodes numériques intensives dont la complexité computationnelle croît avec la sophistication des produits et la précision des modèles. Cette sophistication répond

à deux impératifs concurrentiels majeurs, d'une part l'adaptation précise aux besoins des clients et, d'autre part, la performance des modèles de gestion de risque, déterminante pour la compétitivité de l'activité.

PRINCIPES ARCHITECTURAUX

Face à ces défis, l'architecture des systèmes de calcul doit s'organiser autour de principes structurants. La séparation du fonctionnel et du technique et l'optimisation des ressources via une orchestration capable de rapprocher le calcul de sa donnée afin de minimiser les latences et de conserver des données coûteuses à calculer potentiellement réutilisables.

Pour rendre possible une optimisation à un niveau global, il est nécessaire de se doter d'une abstraction des services de calcul sous forme de tâches, de données et d'interactions entre ceux-ci. La recherche de la cartographie la plus adaptée est une source féconde d'optimisations très matérielles.

Les stratégies d'optimisation doivent s'adapter aux différents contextes d'utilisation. On privilégiera l'optimisation de la durée des traitements récurrents avec exploitation maximale des ressources disponibles ou la rapidité de réponse unitaire pour les traitements temps-réel.

Dans tous les cas, les facteurs limitants restent les ressources de calcul, la mémoire et leurs interactions. L'orchestration doit gérer efficacement plusieurs dizaines de milliers de cœurs répartis sur différents sites et environnements cloud, tout en optimisant une

fonction de coût globale intégrant l'empreinte carbone – un défi technique considérable.

L'ÉMERGENCE DE L'IA

L'intelligence artificielle transforme progressivement l'approche du calcul dans le secteur financier. Plus récemment, l'ensemble des cas d'usage de l'IA générative se répandent dans les différents services de l'industrie financière.

Les modèles d'IA sont eux-mêmes consommateurs intensifs de ressources. Cette approche présente plusieurs caractéristiques distinctives comme l'utilisation d'algorithmes standardisés via des librairies externes, principalement en Python et le recours aux accélérateurs matériels spécialisés (GPU/TPU). Ces caractéristiques renforcent la nécessité de s'appuyer sur une architecture fondée sur l'adaptabilité au contexte d'utilisation.

Au-delà de ses applications métiers, l'IA est une technique performante pour l'optimisation de la distribution globale des calculs elle-même, dans des environnements hybrides incluant le cloud. Elle peut par ailleurs être entraînée à l'apprentissage des fonctions métier les plus coûteuses en calcul pour en réduire drastiquement le temps d'exécution.

L'intégration de l'IA dans un système d'information pose de nouveaux défis. Le contrôle de fonctions qui évoluent par leur utilisation même ou la gestion du périmètre de validité souvent flou de ces fonctions

s'avèrent délicats en pratique. Ces considérations orientent le choix des premiers cas d'usage privilégiés pour l'IA. Elle est particulièrement adaptée au remplacement d'heuristiques pour la résolution de problèmes complexes. En revanche, la prudence s'impose lorsqu'une précision rigoureuse est impérative.

PERSPECTIVES : L'INFORMATIQUE QUANTIQUE

L'informatique quantique apparaît pleine de promesses. Les premières machines opérationnelles permettent déjà d'explorer des applications spécifiques exploitant les propriétés quantiques pour résoudre certaines classes de problèmes avec une efficacité potentiellement révolutionnaire.

Des premiers cas d'usage opérationnels dans l'industrie financière ont vu le jour sur le thème de l'optimisation combinatoire de portefeuille et l'optimisation des fonds propres réglementaires. De nombreuses autres pistes sont en phase d'exploration : par exemple les algorithmes quantiques appliqués à la résolution d'EDP, au clustering ou à l'IA.

Cette technologie transforme d'ores et déjà les approches plus standards avec l'exploitation des réseaux de tenseurs et les méthodes inspirées du quantique. Elle pourrait constituer le prochain paradigme de calcul pour certaines applications financières, complétant l'arsenal des méthodes HPC et IA dans l'écosystème technologique du secteur. ■

LE CLOUD CLÉ DE VOÛTE DE L'INNOVATION, D'AUJOURD'HUI ET DE DEMAIN !



Yaniv Fdida

Chief Product and Technology Officer, OVH Cloud

Les observateurs savent que le numérique fonctionne selon des cycles d'innovation. Ceux-ci redéfinissent généralement le paysage établi. Ainsi à la toute fin des années 90, l'éclosion de l'Internet a entraîné la fameuse vague du « dotcom ». C'est d'ailleurs en 1999 que naît OVH Cloud ! Depuis, il y a eu bien d'autres vagues, et notamment celle des smartphones qui a mis la puissance de calcul des ordinateurs dans nos poches ou sur nos poignets. Un cycle plus discret mais non moins transformatif est étroitement lié à cette démocratisation des usages toujours plus instantanés et nomades : le cloud. Il accompagne en effet le mouvement de dématérialisation des contenus (musique, vidéo, photos, stockage, etc.). Transparent, le cloud est aujourd'hui indispensable à nos modes de consommation.

Plus récemment c'est bien sûr la vague de l'IA qui s'est imposée dans l'actualité et dans notre quotidien. L'irrésistible ascension de chatGPT nous a plongés dans un nouveau cycle dont il est bien difficile aujourd'hui de prévoir l'ampleur des répercussions. Une chose est sûre, l'intelligence artificielle n'est pas un épiphénomène, elle s'inscrit dans le temps long et promet de révolutionner certains domaines qu'il s'agisse du transport, de la santé ou de la robotique. Tout ceci demande bien sûr une vigilance de chaque instant tant la matière première servant à entraîner les intelligences artificielles, à savoir la donnée, est sensible que ce soit pour les individus ou les entreprises. La donnée, devenue plus que jamais or noir à l'ère de l'intelligence artificielle, doit être utilisée de façon responsable et éthique. Il nous appartient en tant que collectif de fixer le curseur sur ce qui paraît acceptable, ou non, dans les applications qui peuvent en découler. Comme à chaque nouvelle évolution technologique majeure, la question des usages est au centre des préoccupations.

À l'avenir, nous pensons que l'informatique quantique constitue la prochaine disruption, permettant d'adresser des problèmes jugés aujourd'hui encore insolubles. Le sujet reste balbutiant mais l'Europe se démarque

déjà par un grand nombre d'acteurs et une incroyable diversité d'approches. Notre fondateur, Octave Klaba, a coutume de dire que le vingtième siècle était atomique, et que le prochain sera quantique. C'est fort de cette vision que nous déployons nos savoir-faire pour connecter au cloud les premiers ordinateurs quantiques afin de préparer les écosystèmes de demain en rendant leur puissance de calcul accessible au plus grand nombre.

COMMENT OVH CLOUD APPROVOISE L'INTELLIGENCE ARTIFICIELLE ?

Depuis 2017, OVH Cloud infuse l'intelligence artificielle à travers ses solutions. En intégrant les principes du machine learning pour ses propres usages internes (solutions de maintenance prédictive, antifraude, etc.), le Groupe s'est bien positionné pour démocratiser l'accès à l'intelligence artificielle. C'est ainsi que sont nés divers services simples et abordables permettant l'accélération de l'acculturation de l'IA par nos clients allant de l'entraînement de modèles à leur déploiement en passant par la mise à disposition de modèles open source sur étagère à destination des développeurs. Et si nous parlions plus haut de l'or noir qu'est la donnée pour l'IA, il est clair pour nous qu'il n'y a pas d'IA éthique sans souveraineté de la donnée. C'est dans ce but précis que nous avons développé tout un ensemble de services pour stocker, manipuler la donnée, l'ingérer et la transformer. De cette approche unique autour de la donnée découle une proposition de valeur éthique et de confiance pour l'IA.

Une intelligence artificielle, dont les usages impliquent divers impératifs matériels. Les cartes d'accélération qui lui sont dédiées, et particulièrement les GPU que nos amis gamers connaissent bien, sont aujourd'hui indispensables, mais pas que ! La puissance de calcul ou compute repose certes sur un tandem CPU/GPU mais également sur le réseau et le stockage qui se doivent d'être ultra optimisés et à l'état de l'art pour favoriser de faibles temps de latence. Si à l'avenir

• • •

LE CLOUD CLÉ DE VOÛTE DE L'INNOVATION, D'AUJOURD'HUI ET DE DEMAIN !



Yaniv Fdida

Chief Product and Technology Officer OVH Cloud



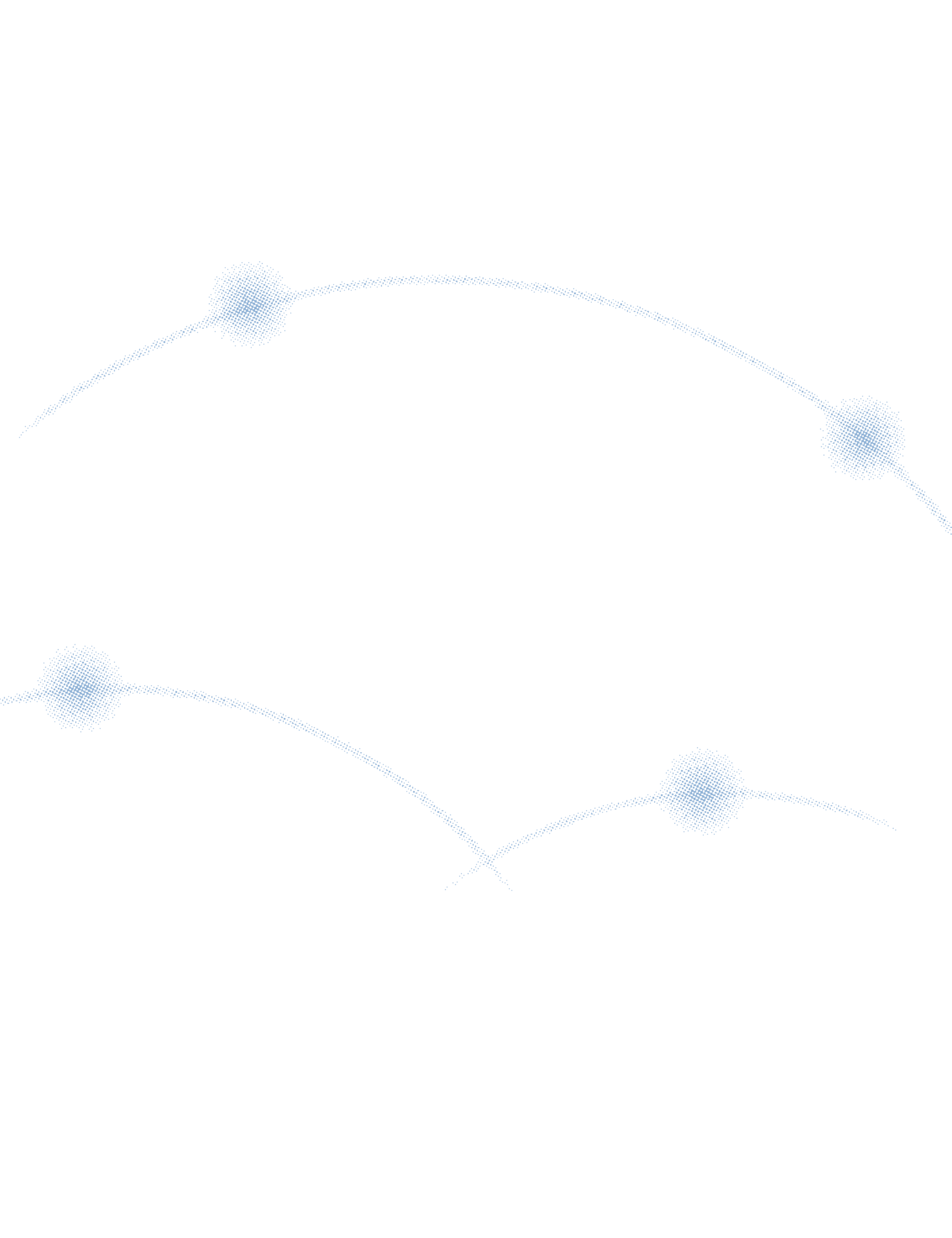
l'intelligence artificielle peut recourir à des accélérateurs dédiés notamment pour l'inférence, solution que nous étudions actuellement, le GPU reste incontournable en cette année 2025. C'est tout aussi excitant que challengeant, ce type de puce étant connu pour sa forte consommation énergétique et son échauffement. En effet, l'impact environnemental des GPU est tout sauf anodin.

Le corollaire c'est la nécessaire adaptation des datacenters et de leur urbanisation pour accueillir ces GPU, les alimenter électriquement et les refroidir avant de distribuer leur puissance de calcul. Depuis 2003, OVHcloud a fait le choix, pionnier pour l'époque, de refroidir ses serveurs avec une boucle d'eau en circuit fermé. Cela réduit la consommation en électricité et en eau des climatisations abondamment utilisées dans les datacenters traditionnels. Nos équipes de R&D ont travaillé sans relâche à l'élaboration de différentes générations de systèmes de refroidissement à eau et ont mis au point des solutions permettant de refroidir jusqu'à 8 GPU par serveur.

CLOUD ET SOUVERAINETÉ EN 2025

L'instabilité géopolitique actuelle confirme la pertinence des choix stratégiques d'OVHcloud et notre positionnement sur un enjeu essentiel : la nécessité pour les États, notamment en Europe, et pour toutes les organisations, de renforcer leur souveraineté numérique. OVHcloud, se positionne depuis ses origines sur des solutions souveraines et ouvertes et est donc au cœur de cette tendance de fond.

La forte dépendance de l'Europe vis-à-vis de solutions principalement américaines peut avoir de lourdes conséquences, notamment sur nos activités économiques, sur notre vie privée et plus globalement sur notre autonomie stratégique. En sécurisant ses infrastructures numériques critiques, en renforçant sa capacité d'innovation et en faisant le choix de promouvoir un cloud ouvert et de confiance, l'Europe disposera d'une carte maîtresse pour affronter les recompositions géopolitiques en cours. Les décideurs, publics et privés, et investisseurs l'ont compris, il convient désormais d'agir. ■







ASPECTS SOCIÉTAUX



INTRODUCTION : ASPECTS SOCIÉTAUX, DU BON ET DU MAUVAIS

Jean-François Prevéraud
Journaliste

Si les technologies numériques sont utilisées de longue date dans quasiment tous les secteurs d'activité, leur impact sur la société a été limité jusqu'à une date récente. Certes, certains emplois ont dû évoluer pour s'adapter au prix de quelques formations. Mais cela allait dans le sens d'un confort de travail accru.

En revanche, l'apparition, puis la démocratisation d'Internet et l'arrivée plus récente de l'Intelligence Artificielle Générative ont changé la donne et impacte toute la société.

UN CHANGEMENT DE MODE DE VIE

Au début des années 90, l'évolution technologique permet l'arrivée d'applications web « grand public », grâce à des moteurs de recherche en ligne performants (Mosaic; Netscape; Yahoo; Lycos; Altavista, Google; Internet Explorer...). Plus besoin d'avoir la connaissance, il suffit de savoir la rechercher. Et l'intégration de l'IA dans les moteurs de recherche facilite grandement les choses, anticipant même les demandes.

Le monde du commerce a été l'un des premiers impactés par la démocratisation des technologies numériques. Des plates-formes comme eBay, Amazon, Alibaba, Rakuten... ont peu à peu accaparé une grande part du commerce mondial, poussant à la faillite nombre d'entreprises tenantes du « Bricks & Mortar ». Et aujourd'hui les services connaissent la même logique d'évolution avec, par exemple, les banques « en ligne », qui poussent à la fermeture des agences bancaires physiques.

Le monde de la communication connaît des révolutions similaires. Déjà le téléphone portable (GSM) a quasiment remplacé le téléphone fixe. En apportant toute la puissance Internet au fond de la poche, il permet à chacun d'accéder, de créer et de diffuser de l'information en temps réel, notamment via les « réseaux sociaux » professionnel (LinkedIn...) ou grand public (Facebook; Instagram; Snapchat; Twitter; WhatsApp; YouTube; TikTok...), avec des applications dérivées comme les jeux et la musique en ligne ou les sites de rencontre.

RESTONS VIGILANTS

Une belle idée que certains détournent pour faire insidieusement du marketing personnalisé ou, pire, diffuser des Fake News (Infox) pour appuyer des idées nauséabondes pour la démocratie. Des réseaux sociaux qui ont aussi un impact indéniable sur les plus fragiles et les jeunes, il n'est de voir les ravages du cyberharcèlement et de la cyberdépendance. À tel point que les états réfléchissent sur les réglementations à mettre en place.

Enfin, n'oublions pas la « fracture numérique » qui s'élargit entre ceux qui ont accès et utilisent les technologies numériques au quotidien et ceux qui ne peuvent pas car cela à un coût trop important pour eux, ou qui n'en ont pas ou plus les capacités.

Reste que ces technologies numériques ont aussi un impact positif sur la société. Il suffit de voir l'impact de l'IA sur l'enseignement, qui modifie les façons d'enseigner et d'apprendre. Une positivité que nos experts vous expliquent dans les pages qui suivent. ■

LA FORMATION FACE À LA VAGUE DIGITALE : NOUVEAUX SAVOIRS POUR NOUVEAUX DÉFIS

Jean-Christophe Jouhaud

Responsable de la Formation, Cerfacs

L'essor rapide de l'Intelligence Artificielle (IA), du Calcul Haute Performance (HPC) et des technologies quantiques transforme profondément les besoins en compétences dans tous les secteurs. Cette mutation s'accompagne d'une pénurie croissante de profils qualifiés, dans le public comme dans le privé. Ce déséquilibre résulte d'une accélération technologique plus rapide que l'adaptation des formations, aggravée par une compétition mondiale intense sur les talents experts.

Pour répondre à cette tension, plusieurs leviers structurants doivent être activés : développement de parcours professionnalisants, co-construction des cursus avec les entreprises, revalorisation des filières scientifiques dès le secondaire, et intégration de modules hybrides IA/HPC/quantique dans les programmes existants.

L'Europe a pris la mesure de cet enjeu stratégique en inscrivant la formation digitale au cœur de plusieurs initiatives majeures : le Digital Europe Programme, EuroHPC JU, EuroQCI, EIT Digital ou encore le réseau ELLIS. L'objectif est clair : bâtir un socle européen de compétences numériques de haut niveau.

La France s'inscrit activement dans cette dynamique, notamment à travers deux projets clés :

- EuroCC2 / CCFR (Teratec, Cerfacs, INRIA Academy, CRIANN et ROMEO), qui fédère plus de 30 centres nationaux de compétences en HPC en Europe ;

- Les EDIH (European Digital Innovation Hubs), qui accompagnent les PME et les acteurs publics dans l'adoption des technologies numériques via la formation, l'expérimentation et l'appui stratégique.

Ces hubs jouent un rôle essentiel dans la formation continue, la sensibilisation sectorielle (santé, agriculture, industrie) et la diffusion territoriale des compétences.

En formation initiale, la France est engagée dans le programme EUMaster4HPC, un master européen d'excellence porté par plusieurs établissements (Sorbonne Université, Université de Strasbourg) et des partenaires en lien avec le monde industriel tels que le Cerfacs par exemple. Ce cursus vise à former une nouvelle génération de spécialistes, capables d'articuler expertises scientifique, technique et applicative autour du HPC.

Pour relever le défi du volume, de la spécialisation et de l'agilité, les nouvelles modalités pédagogiques sont indispensables. L'enseignement à distance joue un rôle crucial :

- Les MOOC initient un large public aux fondamentaux ;
- Les SPOC offrent des parcours approfondis et encadrés ;
- Les formations hybrides allient contenus numériques, ateliers pratiques et accès distant à des infrastructures comme les clusters HPC ou les plateformes quantiques.

Former mieux, plus vite et plus largement est désormais une condition incontournable de compétitivité et de souveraineté technologique. ■

DES ENJEUX INÉDITS DANS UN MONDE EN PROFONDE MUTATION

Thierry Damerval

Envoyé spécial pour la science, la technologie et l'innovation
Ministère de l'Europe et des affaires étrangères
Ministère de l'Enseignement supérieur et de la recherche

2005-2025 : QUE DE BOULEVERSEMENTS DANS CE MONDE EN PROFONDE ÉVOLUTION !

Rappelons-nous qu'il y a vingt ans, les discussions en Europe portaient sur l'économie de la connaissance et l'objectif de consacrer 3 % du PIB à la recherche et développement. Aujourd'hui, c'est l'économie de la défense et la part du PIB qui doit y être consacrée qui sont à l'ordre du jour. Le monde scientifique lui-même a connu de profondes évolutions. Jusqu'au début des années 2000, le monde scientifique a été marqué par une certaine stabilité en termes de classement des nations depuis la création de l'ISI (Institute for scientific information) en 1957. Les États-Unis étaient de loin la première nation scientifique et quatre pays anglophones (États-Unis, Royaume-Uni, Canada, Australie) représentaient plus de 40 % de la production scientifique mondiale et l'accueil de la majorité des étudiants et post-doctorants du monde entier.

Cette cartographie mondiale de la science a profondément évolué au cours des vingt dernières années, générant de nouveaux équilibres et l'émergence de nouveaux acteurs. La part des États-Unis dans la production bibliographique mondiale est passée de 28 % à 14 %. La part de la Chine représente aujourd'hui 20 %, contre 3,9 % en 2000, passant de la 6^e à la 1^{re} place mondiale. L'Inde est passée de la 13^e à la 3^e place. Ces indicateurs strictement quantitatifs sont évidemment l'objet de discussions et de critiques. Se pose notamment la question de la croissance non régulée

du nombre annuel de publications, passé de 800 000 en 2000 à près de 3 000 000 en 2023, un taux de croissance jamais observé dans l'histoire des sciences. Cette croissance est bien supérieure à la croissance du nombre de chercheurs et chercheuses dans le monde, est associée à une course à la publication et pose de nombreuses questions en matière de conduite et d'évaluation des recherches. On constate également au niveau mondial une croissance des problèmes de qualité, de reproductibilité, de fraudes. Le taux de rétractation d'articles à la suite d'erreurs ou de fraudes avérées a ainsi été multiplié par 10 en 20 ans.

Dans ce contexte évolutif et dont l'avenir est difficilement prévisible (les élections présidentielles américaines nous l'ont brutalement rappelé), on peut retenir deux éléments qui ont marqué les deux dernières décennies par leurs enjeux majeurs, tant sur les plans scientifique, technologique, économique que diplomatique :

- L'agenda 2030 des Nations Unies et les 17 objectifs de développement durable adoptés en 2015 : après des rapports fondateurs (rapport Meadows en 1972, rapport Brundtland en 1987 et premier rapport du GIEC en 1990), la prise de conscience collective et le passage à l'action ont été lents.
- L'essor extrêmement rapide des technologies de l'information et de la communication et des sciences et technologies associées (calcul à haute performance, gestion des données, intelligence artificielle, technologies quantiques).

Les applications déjà effectives et les potentialités de l'IA sont considérables. Elles touchent tous les secteurs de la vie sociale et économique, fait rarissime dans l'histoire des sciences et des technologies. Le domaine de la santé en sera sans aucun doute l'un des grands bénéficiaires. Elles modifieront l'enseignement, et concernent tous les domaines de la recherche, les activités de veille, la conception de projets de recherche, voire leur évaluation, la réalisation de synthèses et l'expertise scientifique. Se posent toutefois au niveau international de nombreuses questions sur la régulation des usages, dans un contexte où le poids des acteurs privés l'emporte sur la sphère publique et où la rapidité des développements génère des utilisations avant même que le cadre des régulations soit défini.

UNE TECHNOLOGIE ÉMINEMMENT GÉOPOLITIQUE.

Vladimir Poutine déclarait en 2017 « celui qui dominera l'IA dominera le monde ». En 2020, le think tank américain « Brookings Institution » allait dans le même sens « celui qui sera le leader de l'intelligence artificielle en 2030 dominera le monde jusqu'en 2100 ». La Chine s'est fixée comme objectif de posséder l'ensemble de la chaîne de valeur de l'IA sur son territoire d'ici 2030.

L'IA est d'abord américaine avec un rôle et un poids considérable des grands acteurs privés (les GAFAM), de nouveaux acteurs comme OpenAI, les volumes de données qui ont facilité le deep learning, les infrastructures matérielles. Les enjeux géostratégiques sont évidents lorsque l'on pense que le principal fabricant de puces et fournisseur de l'entreprise américaine Nvidia est la société taïwanaise TSMC. La Chine est évidemment l'autre puissance dominante, avec une croissance extrêmement rapide du volume de données maîtrisées, et dont la plupart sont fermés.

La France, comme l'Allemagne, le Canada, le Japon, la Corée du Sud constituent un groupe de puissances intermédiaires, dont aucune ne pourra prétendre seule jouer un rôle dominant.

Les enjeux et les opportunités se jouent pour nous dans un cadre européen.

QUELLE RÉGULATION INTERNATIONALE ?

En matière de régulation des usages, il est intéressant de faire un parallèle avec une autre grande révolution scientifique : l'émergence du génie génétique et ses applications dans le domaine biomédical. Devant les applications potentielles de ces nouvelles technologies, la communauté scientifique mondiale avait exprimé ses interrogations lors de la conférence d'Asilomar en 1975. Cette mobilisation avait conduit à la création du Comité international de bioéthique de l'Unesco et, en France, à la création en 1983 du Comité consultatif national d'éthique pour les sciences de la vie et de la santé. Les années 1980 connurent de nombreux débats et réflexions associant les différentes parties prenantes, au niveau national, européen et international. Un cadre et des limites furent fixés, comme l'absence d'intervention sur les lignées germinales ou le bannissement du clonage, cadre respecté jusqu'à présent au niveau international.

Une conférence d'Asilomar sur l'intelligence artificielle a bien eu lieu en 2017, mais elle n'a pas eu le même impact que celle de 1975. Au niveau européen, l'AI Act fixe clairement un cadre. Permettra-t-il d'accorder davantage de confiance aux modèles qui y sont développés et contribuera-t-il à une stabilité pour les développeurs et clients des systèmes d'IA ? Ces enjeux sont majeurs pour l'Europe.

Lors du sommet pour l'action sur l'IA, qui s'est tenu à Paris en février 2025, coprésidé par la France et l'Inde, la déclaration pour une IA de confiance, durable et inclusive a été signée par 64 États, dont l'Inde, la Chine, la Canada, l'Australie, l'UE et ses états membres, et l'Union africaine. Il est toutefois significatif que ni les États-Unis, ni le Royaume-Uni ne s'y soient associés.

L'IA et les technologies associées modifient le monde et notre vie quotidienne, avec des bénéfices incontestables mais également des risques dont la maîtrise constituera un défi majeur pour les vingt prochaines années! ■

ENCADRER L'INNOVATION SANS LA FREINER : L'APPROCHE EUROPÉENNE

Alexandru Mateescu

Doctorant, Université Paris1 Panthéon-Sorbonne

L Le développement rapide des technologies numériques - calcul intensif; simulation; intelligence artificielle; et bientôt quantique - transforme nos sociétés. Ces outils apportent des avancées spectaculaires dans la santé, l'éducation, l'industrie et l'environnement, etc., tout en soulevant des risques majeurs: désinformation; fracture numérique; dépendance technologique; cybersécurité, etc. Face à ces défis, la question n'est pas de freiner l'innovation, mais de l'encadrer avec discernement. L'Europe a choisi une voie singulière, exigeante et ouverte, où la régulation devient un levier pour construire la confiance, protéger les citoyens et soutenir la compétitivité. Cette tension entre protection et dynamisme est au cœur du débat contemporain.

ENCADRER SANS FREINER : UN DÉFI COMPLEXE

Encadrer sans freiner l'innovation est un exercice délicat et permanent. À mesure que les technologies progressent, leurs usages se diversifient et se complexifient. Lutter contre la désinformation ne doit pas restreindre la liberté d'expression. Rendre les algorithmes plus transparents ne doit pas décourager la recherche ni compromettre la confidentialité des secrets industriels. Imposer des obligations strictes en matière de cybersécurité ne doit pas paralyser les petites structures qui innovent. Chaque décision implique un arbitrage subtil entre objectifs légitimes, souvent en tension.

Ces tensions se traduisent dans la régulation par une série de choix difficiles. Les règles doivent être suffisamment claires pour assurer la protection des citoyens,

mais assez souples pour laisser place à la créativité et à la compétitivité. Ce n'est pas seulement le contenu des textes qui compte, mais le processus complet qui les fait vivre: leur élaboration, leur mise en œuvre et leur appropriation par les acteurs de terrain.

Cet équilibre ne se construit pas uniquement dans les instances législatives, il prend corps dans des lieux où l'innovation et la régulation se rencontrent. Au moins trois types d'écosystèmes jouent ici un rôle complémentaire:

- Les réseaux d'entreprises, en France et en Europe, qui fédèrent autour de la simulation, du calcul intensif et de l'intelligence artificielle. Ces réseaux, comme Teratec en France, jouent un rôle clé pour favoriser l'expérimentation, la diffusion des technologies et la formation, tout en créant des synergies entre fournisseurs, utilisateurs et laboratoires.
- Du côté académique, de nombreuses structures européennes réunissent des chercheurs issus de disciplines variées — droit, informatique, philosophie, économie, arts — pour explorer l'impact du numérique et notamment de l'IA sur la société et contribuer au développement des savoir-faire et à une compréhension critique des usages.
- Enfin, au niveau européen, la DG Connect et l'initiative EuroHPC mutualisent les ressources de calcul haute performance entre États membres et animent, en liaison avec les États membres, la politique de recherche et développement. Ces actions ont pour objectif que

l'Europe conserve une souveraineté stratégique sur ses infrastructures numériques, tout en soutenant l'innovation et la compétitivité industrielle.

En articulant ces niveaux – entreprise, université, gouvernance européenne et autres – l'Europe transforme la régulation en un processus vivant, où la vision politique, la recherche et l'innovation industrielle se renforcent mutuellement et préparent le terrain aux solutions législatives.

LES SOLUTIONS EUROPÉENNES

L'Union européenne a progressivement construit un cadre réglementaire cohérent pour trouver un équilibre délicat : protéger les citoyens ; soutenir l'innovation ; préserver la maîtrise collective du numérique.

La première étape a été le RGPD (Règlement Général sur la Protection des Données) en 2018. Non seulement il a renforcé la confiance en donnant à chaque citoyen le contrôle de ses données personnelles, mais il a aussi encouragé le développement de services numériques sensibles, comme la télémédecine. L'Europe a ainsi posé un socle commun qui, tout en protégeant la vie privée, a permis l'émergence d'un marché numérique intégré et dynamique.

En 2022, l'Europe a franchi une nouvelle étape avec le DSA (Digital Services Act) et le DMA (Digital Markets Act). Le DSA impose aux plateformes de gérer activement les risques systémiques, tout en encadrant ces actions par des règles de transparence et de recours. Le DMA, complémentaire, réduit la dépendance à quelques géants du numérique tout en ouvrant les marchés à de nouveaux acteurs, stimulant la concurrence et la créativité.

En 2023, le Data Governance Act et le Data Act ont abordé la question cruciale de la circulation des données. Ils ne se contentent pas de sécuriser les flux, ils favorisent une création de valeur partagée. Ainsi, une PME peut non seulement accéder aux données produites par ses machines connectées, autrefois verrouillées par un

fournisseur unique, mais aussi les utiliser pour optimiser sa production et créer de nouveaux services. En protégeant les données sensibles, tout en élargissant l'accès, l'Europe renforce l'autonomie stratégique et la compétitivité de ses acteurs économiques.

En 2024, l'AI Act a complété cet ensemble avec un cadre inédit pour l'intelligence artificielle. Il introduit une logique de proportionnalité :

- Les usages critiques, comme les logiciels médicaux ou les outils de recrutement, sont soumis à des obligations strictes de transparence et de robustesse ;
- Tandis que les usages courants, comme la recommandation de contenus culturels, bénéficient d'un cadre plus souple, qui favorise l'adoption et l'innovation.

Ainsi, l'Europe encadre les IA à haut risque tout en permettant la créativité et la diffusion rapide des solutions les moins sensibles.

Ces quatre piliers – vie privée ; sécurité en ligne ; ouverture des marchés ; encadrement des IA – montrent qu'une régulation bien conçue peut protéger sans paralyser, guider sans contraindre. En articulant protection individuelle, dynamisme économique et souveraineté collective, l'Europe trace une voie singulière, où la régulation devient un levier d'innovation autant qu'un garant des libertés.

BÂTIR UN FUTUR NUMÉRIQUE RESPONSABLE ET DURABLE

En reliant ses efforts autour de ces quatre piliers, l'Europe avance pour montrer qu'il est possible d'innover sans perdre la maîtrise collective du numérique. Cette approche transforme la régulation en un processus vivant, reliant institutions, entreprises et chercheurs dans une dynamique commune. Elle constitue non seulement un modèle pour un futur numérique responsable et durable, mais aussi une invitation à poursuivre le dialogue entre innovation, confiance et souveraineté face aux défis à venir. ■

LE MONDE DE L'ENSEIGNEMENT EST MOTEUR SUR LES NOUVELLES TECHNOLOGIES NUMÉRIQUES



Guillaume Gellé
Professeur des Universités

Pour traiter des aspects formation liés aux nouvelles technologies numériques, nous avons rencontré Guillaume Gellé, Professeur des Universités, ancien Président de l'Université de Reims Champagne-Ardenne, ancien Président de France Universités, ancien membre du CA de Gencî et ancien président du CINES. Il nous a notamment éclairés sur l'impact de l'IA, tant sur les étudiants que sur le corps professoral, ainsi que sur la pédagogie et les cursus de formation.

Parallèlement à ma carrière universitaire, j'ai occupé des postes m'offrant une vision globale du numérique dans le monde de l'enseignement. J'ai récemment rendu au Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation, un rapport sur les opérateurs du numérique en France, qui a débouché début septembre 2025 sur l'annonce de la fusion du CINES, de l'ABES et l'AMUE*, pour renforcer la cohérence et la qualité de l'offre de services numériques, au bénéfice des établissements d'enseignement supérieur et de recherche et de leurs usagers. Mais également à adapter la stratégie française au contexte européen.

L'ENSEIGNEMENT S'ADAPTE À L'IA

L'impact des technologies numériques sur la formation des jeunes et la place croissante prise par l'IA, nécessite bien sûr une adaptation de la pédagogie, mais aussi de faire prendre conscience aux jeunes qu'il s'agit d'un monde en mutation perpétuelle, remplis de pièges, qu'il leur faut maîtriser. Donc leur apprendre dès le plus jeune âge à ne pas croire ce qu'ils voient sur Internet. L'esprit critique est un sujet majeur dans la construction de leur pensée.

La prise de conscience de ces enjeux est maintenant générale et cette révolution de la formation au numérique est engagée. Cela suppose des plans de formation massifs des enseignants avec des formateurs spécialisés, pas forcément issus de nos établissements.

Comme pour toutes les autres disciplines, cet effort de formation au numérique et à l'IA devra commencer dès le plus jeune âge et se poursuivre tout au long de la scolarité. Mais le numérique évolue très vite, c'est un challenge dont la communauté éducative s'est saisie. Par contre,

comme dans toutes les autres disciplines scientifiques, les fondamentaux sont assez statiques. Ce sont ceux qu'on enseigne aux plus jeunes. Il faudra aussi préparer nos élèves à affronter des évolutions qu'on ne connaît pas encore, mais qui impacteront la société à l'avenir.

FORMER DES SPÉCIALISTES

À côté de cette formation généraliste, «le numérique et l'IA pour tous», il va falloir aussi développer des formations de pointe, «le numérique et l'IA pour les spécialistes». Nous avons tous les atouts pour y parvenir : l'une des meilleures écoles de mathématiques ; de grandes écoles d'ingénieurs ; des chercheurs universitaires de très haut niveau ; des financements publics notamment via France 2030. Ces formations de très haut niveau de spécialistes de l'IA sont en place, sur les technologies de pointe et les ruptures technologiques que l'on envisage. C'est aussi, une course contre la montre, non plus technologique, mais internationale avec la Chine ou les États-Unis, qui forment à tout va, ainsi qu'avec des entreprises internationales qui cherchent à débaucher nos meilleurs talents.

Cette stratégie de formation doit s'appuyer sur quelques centres d'excellence, mais aussi sur l'ensemble de nos universités et écoles d'ingénieurs pour assurer une diffusion de la connaissance à tous nos étudiants sur tout le territoire. Il faut donc poursuivre le travail, mené notamment par Gencî, via des programmes comme MesoNet qui à partir de centres d'excellence essaient jusqu'aux universités de territoire. Il faut aussi développer la stratégie européenne d'alliances d'universités et de développement de pôles spécialisés en IA. C'est notre seule chance pour ne pas être distancé par la Chine qui avance très vite.

FORMER MASSIVEMENT TOUS LES INGÉNIEURS

L'IA devient l'un des savoirs fondamentaux de nombreux métiers dont ceux de l'ingénierie. Tous nos ingénieurs généralistes doivent donc être formés correctement dans le domaine de l'IA, pour en saisir le fonctionnement et avoir, a minima, une capacité de petits développements. Un défi immense car si on forme déjà près de

LE MONDE DE L'ENSEIGNEMENT EST MOTEUR SUR LES NOUVELLES TECHNOLOGIES NUMÉRIQUES



Guillaume Gellé
Professeur des Universités

40 000 ingénieurs par an en France, il faudrait en former 80 000 en 2030 pour atteindre les objectifs fixés dans le cadre de la réindustrialisation du pays.

Je ne suis pas trop inquiet sur notre capacité à former, car on a un vivier croissant d'enseignants-chercheurs qui se spécialisent dans l'IA et que l'on va pouvoir mobiliser plus. Et c'est aussi vrai dans les organismes nationaux de recherche (CNRS, INRIA...) où de nombreux chercheurs travaillent avec un sens du service public extrêmement fort. Si notre pays a besoin de formateurs, ils répondront présents face aux grands enjeux nationaux. Restera à trouver les moyens matériels de calcul dont ils auront besoin et à relever le défi de leur alimentation en énergie, car il n'y a pas d'IA sans énergie.

UNE COOPÉRATION QUI DOIT S'ÉTENDRE À L'AFRIQUE

Nous avons des liens forts, souvent basés sur des relations humaines privilégiées, avec nombre d'universités africaines, dont nous avons formé les professeurs aux meilleurs standards européens. Elles sont un vivier de jeunes étudiants motivés et brillants, très demandeurs vis-à-vis de l'IA. Dotés d'un bagage mathématique conséquent, ils deviennent d'excellents développeurs ou ingénieurs.

C'est d'autant plus important que l'IA constitue d'une part, une réponse aux défis économiques, sociaux et environnementaux majeurs du continent et d'autre part parce qu'en période de faiblesse de nos relations diplomatiques, les relations partenariales entre universités constituent un soft-power dont on ne peut se passer.

DES ÉTUDIANTS MOTIVÉS

Si les étudiants se sont rués sur ChatGPT lors de sa sortie pour réaliser plus rapidement leurs travaux et évaluations non surveillées, ils ont très vite compris que l'IA Générative allait avoir un impact sur toute la société et qu'il était de leur intérêt de comprendre comment cela fonctionnait pour l'utiliser au mieux. La démarche a été similaire dans le corps professoral pour : l'utiliser dans la pratique professionnelle ; adapter la pédagogie et les modalités d'évaluation des étudiants ; l'intégrer dans

leur enseignement. Étudiants comme professeurs franchissent ce cap. Ce qui est certain, on n'enseignera et on n'apprendra plus de la même manière avec l'arrivée de l'IA. Profitons-en pour personnaliser l'enseignement avec de vrais outils d'accompagnement des élèves là où ils en ont réellement besoin.

DES INDUSTRIELS DEMANDEURS

L'université, avec l'aide des Régions, de l'État et de l'Europe, ouvre ses centres de calcul, aux entreprises, souvent des PME qui n'ont ni les capacités d'investissement, ni les équipes pour disposer des solutions numériques dont elles auraient besoin. Si cette mutualisation des infrastructures à l'échelle des territoires est une première étape, la formation par apprentissage en est une autre, qui permet à des élèves-ingénieurs de travailler sur des projets de R&D correspondant aux besoins des entreprises locales.

C'est différent avec les grands industriels qui se sont dotés d'infrastructures et d'équipes capables d'utiliser ces nouvelles technologies pour leurs propres besoins. Il me semble nécessaire qu'ils contribuent aussi à la formation de nos enseignants aux métiers qu'ils proposeront à nos étudiants diplômés.

PROCHAIN DÉFI, LE QUANTIQUE

Notre pays est très bien positionné dans le domaine de la recherche sur les technologies quantiques et cela a permis l'émergence de start-up parmi les plus prometteuses au niveau mondial. L'écosystème se structure aussi bien en France qu'en Europe sous l'impulsion de Genci et Euro-HPC. Mais nos universités ne sont pas en reste et développent des programmes de formation de haut niveau via des simulateurs de machines quantiques. Ainsi, le centre de calcul Roméo de l'université de Reims s'est doté d'un simulateur dès 2021 et organise régulièrement des conférences et formations dans le domaine à destination des étudiants et des entreprises et coordonne le volet quantique de MesoNet ■.

* CINES (Centre Informatique National de l'Enseignement Supérieur - ABES (Agence Bibliographique de l'Enseignement Supérieur)
AMUE (Agence de Mutualisation des Universités et Établissements)

L'IA, ENTRE INNOVATION ET RESPONSABILITÉ, LE SOMMET DE PARIS A MARQUÉ UNE ÉTAPE HISTORIQUE

Anne Bouverot

Envoyée Spéciale du Président de la République pour le Sommet pour l'Action sur l'IA,
Co-Présidente du Conseil de l'IA et du Numérique,
Présidente du Conseil d'Administration de l'Ecole Normale Supérieure.

« DANS LA VIE, IL N'Y A RIEN À CRAINDRE. TOUT DOIT ÊTRE COMPRIS. ».

Ce sont les mots de la grande scientifique, Marie Curie, rappelant que le progrès découle des échanges d'idées, de cultures et de visions du monde. C'est dans cet esprit que s'est tenu à Paris en février 2025, au Grand Palais le Sommet pour l'Action sur l'IA (« AI Action Summit ») six mois après les Jeux Olympiques et Paralympiques de 2024 et quinze mois après le premier sommet IA de Bletchley Park au Royaume-Uni.

Un Sommet pour ouvrir l'IA au monde, avec en son centre Paris et l'Europe, accueilli par la France et coprésidé par le Président Macron et le Premier Ministre Indien Modi.

L'objectif principal du sommet était de réconcilier science, standards et solutions. Le Grand Palais a été le témoin d'exemples concrets, d'innovation et d'annonces d'investissement pour une IA digne de confiance et d'intérêt général. Nous avons souhaité aller au-delà de la perspective « science-fiction » de l'IA pour démontrer ses applications tangibles et tenter de réconcilier la transition numérique avec la transition écologique, tout en veillant à ce que l'IA augmente les emplois plutôt que de les remplacer.

40 000 PARTICIPANTS

Plus de 800 contributeurs issus de plus de 100 pays, provenant de gouvernements, d'organisations internationales, du monde universitaire, de l'industrie et de

la société civile, ont travaillé avec nous pour préparer et façonner ce sommet. Près de 40 000 personnes ont participé à plus de 250 événements labellisés en France et à l'étranger entre juillet 2024 et février 2025, en amont de la plus grande rencontre mondiale sur l'IA jamais organisée à Paris. Le sommet a réuni 1500 participants du monde entier avec 148 intervenants de 43 pays avec une attention particulière portée à l'égalité de genre, comme en témoignent les 48 % de femmes parmi les intervenants.

Outre le Président Macron et le Premier Ministre Modi, le Sommet a réuni le Vice-Président US JD Vance, la Présidente de la Commission Européenne von der Leyen, le Vice Premier Ministre chinois Zhang Guoqing, la « marraine de l'IA » Fei Fei Li, les prix Nobel Demis Hassabis et Daron Acemoglu, les prix Turing Yann Le Cun et Yoshua Bengio, les dirigeants d'entreprise Sam Altman (OpenAI), Arthur Mensch (Mistral), Gundbert Scherf (Helsing), Choi Soo-Yeon (Naver), Sundar Pichai (Microsoft), Martin Kon (Cohere), Dario Amodei (Anthropic), et bien d'autres.

UNE FEUILLE DE ROUTE

Plus de 100 actions et engagements concrets ont été annoncés pour favoriser une IA digne de confiance, accessible à tous et servant l'intérêt public. Nous les avons appelés les « Actions de Paris pour l'Intelligence Artificielle », une feuille de route pour le développement de l'IA basé sur le partage des sciences, des solutions et des normes communes.

Parmi les initiatives majeures, nous avons lancé :

- Current AI : une fondation internationale pour l'IA d'intérêt public, dotée d'un investissement initial de 400 millions d'euros sur cinq ans, et soutenue par dix pays dont la France, l'Inde, l'Allemagne et le Kenya. Cette fondation se concentrera sur les données, l'ouverture et la responsabilité des modèles d'IA.
- La Coalition pour une IA Durable, initiée par la France, le Programme des Nations Unies pour l'environnement (PNUE) et l'Union Internationale des Télécommunications (UIT), afin de réduire les coûts énergétiques et environnementaux de l'IA et de contribuer à atteindre les objectifs de développement durable des Nations Unies.
- Le Pacte pour une IA Digne de Confiance dans le Monde du Travail, signé par près de 60 multinationales, visant à promouvoir le dialogue social, la sécurité et la santé au travail, ainsi que la productivité et l'inclusivité.
- L'Institut National Français d'Évaluation et de Sécurité de l'IA (INESIA), pour analyser les risques systémiques de sécurité nationale et évaluer les performances et la sécurité des modèles d'IA, notamment dans la lutte contre la désinformation.
- Des engagements concernant le maintien du contrôle humain dans les systèmes d'armes basés sur l'IA et la pose des bases d'un « dialogue mondial » et d'un « panel scientifique » pour renforcer la gouvernance

internationale. Plus de 60 pays ont signé une déclaration politique sur l'IA, forgeant une vision commune pour une IA digne de confiance, durable et inclusive.

330 MILLIARD D'EUROS

Des investissements sans précédent ont été annoncés pour positionner la France et l'Europe à l'avant-garde de l'IA :

- 109 milliards d'euros d'investissements privés en infrastructures d'IA en France ;
- 200 milliards d'euros d'investissements publics et privés en Europe via l'initiative InvestAI et l'initiative EU AI Champions, qui a réuni 120 grandes entreprises et start-up européennes ;
- 20 milliards d'euros d'investissements de la Commission européenne dans l'infrastructure pour l'IA (AI GigaFactories).

Ce Sommet a marqué un tournant pour les futures rencontres internationales sur l'IA : celui de l'ouverture, de l'action et du progrès partagé pour développer des IA dans l'intérêt public.

La diplomatie de l'IA est également scientifique et culturelle, et le système multilatéral est la clé de voûte de cet édifice, que je compte continuer à soutenir dans les futurs événements sur l'IA comme le Sommet pour l'Impact à Delhi en février prochain. ■

..... MENTIONS LÉGALES

« Les Clés du numérique » est édité par Teratec
Campus Teratec - 2, rue de la Piquetterie
91680 BRUYÈRES-LE-CHATEL - France
infos@teratec.fr - www.teratec.eu

..... PUBLICITÉ ET RELATIONS EXTÉRIEURES

Jean-Pascal Jégu - jean-pascal.jegu@teratec.fr

..... DIRECTEUR DE LA PUBLICATION

Didier Besnard, président, Teratec


..... COMITÉ DE RÉDACTION

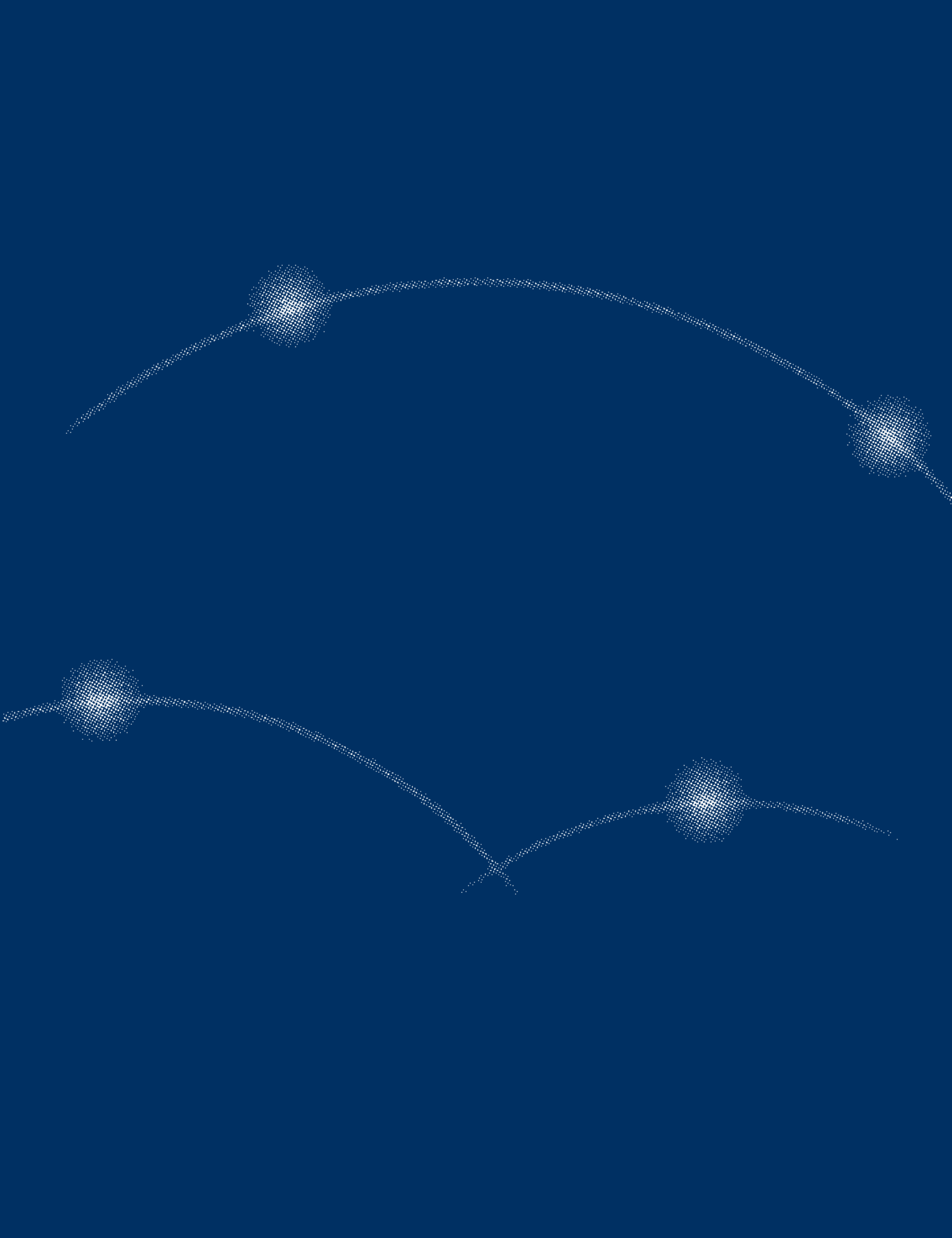
Guillaume Colin de Verdière, CEA
Hervé Mouren, Teratec
Jean-Pascal Jégu, Teratec
Jean-François Prevéraud, Journaliste
Christian Saguez, Teratec
Emmanuelle Vergnaud, Teratec

..... CRÉATION / IMPRESSION

Maindor Digital - o.rouquet@maindor.com

Dépôt légal : octobre 2025





LES CLÉS DU NUMÉRIQUE

Le développement rapide des technologies numériques bouleverse profondément nos façons de travailler et de concevoir produits et services. En quelques années le numérique - la simulation et le calcul intensif et maintenant l'IA et le quantique - s'est imposé comme une composante essentielle de notre développement économique. Il est devenu un des principaux différenciateurs de compétitivité et d'efficacité des entreprises et, en même temps, un enjeu majeur de souveraineté.

L'objectif de Teratec est d'assurer à ses membres la maîtrise de ces technologies et de faciliter leur diffusion et leur déploiement. La formation, initiale et continue, en sera une composante essentielle. Les entreprises qui investissent le plus dans ces technologies bénéficieront d'un avantage compétitif déterminant.

Ceux qui les maîtriseront le mieux seront les leaders de demain.

Hervé Mouren
Directeur, Teratec