

“Nous ne savons pas jusqu’où peut aller l’IA” : les dessous de Kyutai avec Patrick Pérez, CEO

GRAND ENTRETIEN - Lancé en novembre 2023, Kyutai est l’un des labos de recherche en intelligence artificielle les plus en vue. Il est notamment à l’origine de Moshi, une IA générative texte-voix.

Temps de lecture : minute

15 janvier 2025

Moins d’un an après son lancement, Kyutai dévoile Moshi, une intelligence artificielle générative multimodale. Elle est capable de générer du texte mais aussi de discuter vocalement avec un humain. Elle reconnaît jusqu’à 70 émotions. Une première mondiale. Lancé en novembre 2023 par trois figures de la tech et de l’économie, Xavier Niel, fondateur d’Iliad, investisseurs, à l’origine notamment de l’école 42, d’Hectar... Rodolphe Saadé, PDG du groupe CMA-CGM, et Eric Schmidt, ancien directeur général de Google, Kyutai a tout suite suscité des attentes fortes de la part de la communauté IA. Des attentes qui n’ont pas été déçues.

En ce début d’année, le laboratoire livre un “preview” d’Helium-1 : une première version d’un modèle d’IA textuelle et multilingue (français, anglais, allemand, italien, espagnol et portugais pour l’instant) sur lequel travaille actuellement Kyutai dans le cadre de recherches sur des modèles d’IA plus sobres et pouvant tourner en local : sur ordinateur ou téléphone. Ce modèle s’appuie sur deux milliards de paramètres.

Aujourd’hui composé d’une quinzaine de collaborateurs, l’équipe est composée de chercheurs passés par les meilleurs laboratoires publics et privés : Meta, Apple, Google, Polytechnique ou encore l’ENS et est dirigée

par Patrick Pérez. Ce scientifique, spécialiste de la “computer vision”, a exercé à l’INRIA, chez Microsoft avant de créer le laboratoire de Valéo.

Pour *Maddyness*, il dévoile certains projets de recherche de Kyutai et revient sur l’importance de ce type de laboratoire dans la course effrénée à l’intelligence artificielle.

M: Comment définiriez-vous votre manière de faire de la recherche chez Kyutai ?

P.P. : Il ne s'agit pas de recherche théorique. Nous menons ce que l’on appelle de la recherche "amont" : elle n’a pas pour objectif d’être directement finalisée pour des applications spécifiques, des domaines précis, des verticales ou des filières. Toutefois, l’une de nos missions principales consiste à produire une recherche véritablement utile, à concevoir des objets de recherche qui servent non seulement la communauté scientifique, mais également les développeurs et les entreprises, qu’il s’agisse de startups ou de grands groupes.

Notre travail porte sur des intelligences artificielles conçues pour être utiles, tout en restant généralistes et transversales. C’est une recherche résolument appliquée, intégrant une forte composante d’ingénierie.

L’open-source : valeur capitale de Kyutai

M: Lors de votre événement de présentation de Moshi, votre modèle d’IA générative textuelle et vocale, vous aviez annoncé mettre vos codes à disposition de tous. Quelle est la place de l’open source dans votre démarche ?

P.P. : Le partage de la recherche que nous menons avec nos différentes communautés constitue une valeur fondamentale du projet Kyutai. Ces communautés incluent les scientifiques, les développeurs et les

industriels. Il est essentiel, pour des raisons presque politiques, qu'existent des acteurs de premier plan dans la recherche ouverte en intelligence artificielle, particulièrement sur les grands modèles.

Nous partageons nos travaux afin d'éviter que cette technologie ne devienne exclusivement propriétaire, développée et contrôlée uniquement par des groupes privés dictés par leurs agendas commerciaux. Cela représente une priorité pour nous, comme nous l'avons souligné lors de notre événement en juillet dernier, où nous avons annoncé l'accès à Moshi. Depuis cet événement, nous avons tenu nos engagements en partageant une partie des modèles d'IA générative sous-jacents à Moshi sur des plateformes spécialisées comme GitHub ou Hugging Face.

Durant l'été, nous avons consacré de nombreux efforts à la rédaction d'un article scientifique très complet, retraçant la genèse de notre travail, nos choix techniques et les évaluations des modèles, un élément crucial dans le progrès scientifique. Cet article, extrêmement détaillé, compte près de 70 pages. Bien que nous doutions qu'il soit lu en entier par beaucoup, il revêt pour nous une importance capitale.

Notre démarche ne se limite pas au partage des modèles ou du code : nous menons un véritable travail scientifique qui permet aux autres chercheurs de se comparer à nous, de s'inspirer de nos avancées, voire de les surpasser. Cet article, publié en préprint sur la plateforme arXiv, fait partie intégrante de notre mission de chercheurs.

M: Kyutai est née de la volonté de ses trois fondateurs et mécènes : Xavier Niel, Rodolphe Saadé et Eric Schmidt. C'est un laboratoire à but non lucratif, vous fonctionnez uniquement avec les fonds des trois fondateurs?

P.P. : En effet, lors du lancement du laboratoire, trois mécènes fondateurs

ont été publiquement annoncés. Ces derniers ont mobilisé des ressources financières considérables autour de ce projet, ce qui nous permet aujourd'hui de fonctionner pleinement. Cet apport est notamment essentiel pour garantir l'accès à une infrastructure de calcul particulièrement puissante.

M: Quel était le sens de cette démarche qui peut paraître à contre-courant de l'époque actuelle ? Et qu'est-ce qui vous a convaincu de rejoindre ce projet ?

P.P. : Leur idée de départ, bien qu'ils viennent de mondes assez différents, était de rassembler des ressources conséquentes pour permettre l'émergence d'un acteur majeur dans le domaine de l'intelligence artificielle ouverte. Cela leur semblait essentiel à un moment et dans un contexte où disposer d'un nouvel acteur ouvert véritablement crédible était crucial. Les technologies liées à l'IA prennent une importance croissante pour la société, l'économie et la science, et il leur paraissait fondamental que leur développement ne soit pas exclusivement entre les mains d'entreprises privées, souvent engagées dans une compétition féroce et moins enclines à partager avec la communauté.

Cette vision se justifie d'autant plus que la recherche en IA exige des capitaux considérables et des compétences pointues. Les enjeux sont ainsi à la fois sociétaux, géopolitiques et industriels.

Ce qui m'a convaincu de rejoindre ce projet, c'est précisément cette vision ouverte, l'opportunité de me retrouver au cœur de l'action, doté des moyens financiers et humains pour agir. L'équipe, déjà en cours de constitution, rassemblait des talents exceptionnels, tous passés, comme moi, par des expériences dans la tech américaine. Ce projet était incroyablement stimulant et enthousiasmant, avec une dimension presque politique, le tout à Paris et à un moment charnière pour la

discipline.

M. : Est-il question qu'un autre mécène vous rejoigne ?

P.P. : On adorerait ! Ces questions restent tout à fait ouvertes : accueillir d'autres donateurs ou mécènes, qu'ils soient français ou non, est une perspective qui nous intéresse vivement.

Doubler la taille de l'équipe suscite naturellement le désir de disposer de davantage de ressources de calcul. Cela ouvrirait également la voie à des projets encore plus ambitieux.

Vers une IA plus sobre

M. : Quelles sont les grandes orientations de recherche chez Kyutai ?

P.P. : Nous avons encore beaucoup de travail devant nous pour poursuivre le développement de Moshi, notre modèle fondation multimodal texte voix, qui a été au cœur de nos activités durant la première année. Ce projet reste une priorité, avec des avancées à réaliser sur plusieurs fronts : la diversité des langues, l'augmentation des capacités et l'élargissement des modalités.

À ce jour, le modèle repose sur deux modalités principales : le texte et l'audio, ce dernier étant spécifiquement orienté vers la langue parlée. Cependant, nous envisageons d'intégrer d'autres modalités, notamment la vision, qui est d'ailleurs mon domaine d'origine. L'idée que Moshi puisse "avoir des yeux" est une perspective particulièrement captivante sur laquelle nous travaillons. Nous avons d'ailleurs évoqué cette ambition lors de l'événement ai-Pulse organisé en novembre à Station F.

Nous souhaitons également étendre les capacités linguistiques du modèle. Pour l'instant, Moshi est principalement centré sur le français et l'anglais, mais nous sommes très intéressés par l'ajout d'autres langues,

en particulier européennes. Cela soulève des questions importantes concernant la disponibilité et l'accessibilité des données nécessaires.

Un autre axe majeur de nos travaux concerne l'efficacité et la sobriété relative du modèle. Les modèles généralistes que nous développons sont lourds à concevoir et à déployer. Nous cherchons à améliorer leur efficacité, non seulement en réduisant les besoins en données et en ressources de calcul pour leur entraînement, mais également en optimisant leur utilisation.

Par exemple, nous travaillons sur des modèles de plus petite taille, capables de fonctionner localement, c'est-à-dire directement sur un ordinateur ou un téléphone portable, sans nécessiter de connexion Internet. Cette approche présente de nombreux avantages : elle favorise une utilisation plus sobre en énergie, limite la dépendance au cloud, et rend les outils plus faciles à intégrer dans divers environnements. Par ailleurs, cette autonomie est particulièrement utile dans des cas d'usage où la connexion Internet est inexistante ou lorsque des contraintes de sécurité interdisent le recours au cloud.

M. : Jensen Huang, le fondateur et CEO d'Nvidia, a présenté sa vision de l'IA et de son évolution au CES. Il a notamment évoqué l'agentic AI puis la physical AI, qui arriveraient à la suite de l'IA générative. Quelle est votre vision de l'évolution de l'IA?

P.P. : Les étapes évoquées relèvent des jalons classiques de l'évolution de l'intelligence artificielle. Toutefois, il est difficile d'évaluer précisément leur caractère déterminant et les échéances auxquelles elles pourraient être atteintes.

Il est certain que le besoin d'agents IA se fera de plus en plus sentir. Ces formes d'intelligence artificielle, qui trouvent naturellement leur place dans le contexte de l'automatisation du travail, répondront à une

demande croissante. L'IA agentique représente une évolution naturelle, encouragée par les progrès des modèles fondation. Bien qu'il ne s'agisse pas d'une nouveauté, ces avancées permettent désormais d'envisager des applications beaucoup plus larges.

Nous constatons que les modèles d'IA atteignent aujourd'hui des niveaux impressionnants d'intelligence et de capacités. Les progrès réalisés en seulement quelques années sont stupéfiants et ouvrent la voie à des agents intelligents capables d'exécuter des tâches toujours plus complexes, d'une manière de plus en plus coordonnée.

Quant à l'IA physique, qui correspond à la robotique, cette évolution s'inscrit dans une continuité historique. Depuis l'origine, l'intelligence artificielle et la robotique se nourrissent mutuellement. Ces disciplines, bien que distinctes, incarnent différentes facettes d'un même projet : concevoir des systèmes informatiques capables de réaliser des tâches complexes, parfois inaccessibles à l'homme. Dans certains cas, il s'agit d'assister ou d'automatiser des actions humaines pour les rendre plus efficaces et performantes. Dans d'autres, il s'agit d'exécuter des tâches qui dépassent les capacités humaines. Cette progression semble naturelle et logique.

La prospective reste néanmoins un exercice délicat. Prenons l'exemple de la voiture autonome : ayant travaillé dans ce domaine pendant une dizaine d'années, j'ai observé des avancées majeures, notamment grâce aux progrès des modèles de perception, de compréhension et de planification, devenus beaucoup plus puissants.

Cependant, malgré des prédictions optimistes faites il y a quelques années, nous ne sommes pas encore au stade où des voitures totalement autonomes circulent librement dans nos rues. La complexité de ces systèmes, les défis liés à leur fiabilité, leur robustesse et leur acceptation par la société ralentissent leur déploiement. Cela nous invite

à faire preuve de modestie dans nos prévisions.

Il est dans l'ordre des choses que les dirigeants de grands groupes partagent leur vision et présentent des étapes-clés. Toutefois, ces ambitions s'appuient sur les efforts continus de la communauté scientifique et technologique pour atteindre ces objectifs ambitieux.

M. : À votre avis, jusqu'où peut-on aller dans l'IA ? Vous disiez tout à l'heure, on ne sait pas où est-ce qu'on sera dans 5 ans.

P.P. : Non, on ne le sait pas et même d'un point de vue technique. Les outils actuellement utilisés dans le domaine de l'intelligence artificielle, en particulier les IA génératives et les grands modèles de langage, reposent sur des technologies qui, bien qu'impressionnantes, ne sont pas véritablement nouvelles et présentent certaines limites. L'une de ces limites majeures réside dans leur coût : le coût de développement, le besoin massif en données et la taille conséquente des modèles. Il est difficile d'imaginer que ces technologies actuelles marquent la fin de l'histoire ou représentent l'apogée ultime de l'intelligence artificielle. Les réseaux de neurones actuels, bien qu'immenses et complexes à concevoir, ne sont sans doute pas une solution définitive. Atteindre des niveaux de performance équivalents, voire supérieurs à ceux d'aujourd'hui, avec des modèles plus économes en ressources, plus simples à développer et à déployer, est un objectif non seulement souhaitable mais inévitable. Nous nous orienterons naturellement vers ces avancées.

Les interrogations autour des capacités et des compétences de ces systèmes informatiques, mises en parallèle avec l'intelligence humaine, suscitent d'importants débats et nourrissent de nombreux fantasmes. Des surprises sont à prévoir, cela ne fait aucun doute, mais l'avenir nous en dévoilera davantage.

Ce que je trouve particulièrement fascinant, c'est ce que l'évolution actuelle de l'intelligence artificielle révèle sur l'intelligence humaine et animale. Une des choses frappantes avec les grands modèles de langage, qui permettent notamment de concevoir des agents conversationnels, est leur capacité à reproduire, créer et utiliser le langage humain de manière remarquablement convaincante. Cela apporte un éclairage nouveau sur le langage humain lui-même : par le calcul, il est possible de simuler et d'exploiter le langage avec une sophistication impressionnante.

Je pense que c'est là une des grandes leçons de l'émergence des grands modèles de langage, et cela intéresse également ceux qui s'interrogent sur les capacités humaines à parler et à communiquer.

“Il y a d'excellentes universités en France”

M. : Pour revenir sur un sujet un peu plus concret, pour vous, pour les entreprises, quelle est la meilleure manière d'adopter l'IA et comment est-ce qu'on peut aider les industriels à arriver dans cette course ?

P.P. : Encore une fois, si l'on évoque l'intelligence artificielle générative et multimodale, il est indéniable qu'elle transforme profondément les grands groupes, les entreprises et leurs métiers. Ces technologies, qui touchent à l'analyse et à la production de documents, d'images, de textes, et bien d'autres domaines, impactent tous les secteurs reposant sur la manipulation de vastes volumes de données hétérogènes et complexes – ce qui englobe, en réalité, la quasi totalité des industries. Ainsi, il n'est pas surprenant que ces questions préoccupent les entreprises et leurs dirigeants.

Ce qui rend la situation particulière, c'est que ces technologies sont déjà commercialisées par des acteurs de la tech, alors même qu'elles sont encore en pleine évolution. Cette dynamique, relativement nouvelle, peut

être source de stress pour les entreprises et leurs décideurs. Néanmoins, il est probable que certaines choses se stabilisent avec le temps : les prix devraient baisser, et les offres, gagner en clarté.

Un aspect essentiel pour les dirigeants et les décideurs est d'acquérir une compréhension approfondie de ces outils : leurs capacités, leurs limites, les implications de leur déploiement en termes de coûts et de fiabilité. L'acculturation et la pédagogie autour de ces technologies représentent donc des enjeux majeurs. Pour faire face à ces défis, il est crucial que les entreprises soient accompagnées, soit par des partenaires externes compétents, soit en développant ces compétences en interne.

M. : Où est-ce que vous recrutez vos chercheurs ? A votre avis, quelle est la meilleure université ?

P.P. : Il n'y a pas une meilleure université. Il existe d'excellentes universités et formations en France, en Europe et à l'international. En France, cela dépend du profil recherché.

Pour le recrutement, les approches diffèrent selon qu'il s'agisse de chercheurs ou d'ingénieurs. En général, les formations d'ingénieurs en informatique, en mathématiques appliquées et en intelligence artificielle en France, notamment dans les grandes écoles, sont d'un très bon niveau. Ces établissements forment des profils solides, bien que parfois un peu orientés vers la théorie.

L'étape suivante est tout aussi cruciale. Pour le recrutement de chercheurs, nous examinons où ces derniers ont réalisé leur thèse et leurs stages de recherche. Les thèses sont souvent menées au sein de laboratoires académiques, parfois en collaboration avec des industriels. La France et l'Europe comptent de nombreux laboratoires académiques de grande qualité, avec des équipes qui forment d'excellents jeunes chercheurs.

En France, des institutions comme Sorbonne Université, l'INRIA, l'ENS, les Ponts et Chaussées ou Télécom Paris, parmi d'autres, se distinguent. Et cela ne se limite pas à Paris. En Europe, des universités de renom comme Oxford, Cambridge et les institutions londoniennes offrent également des formations et des opportunités de recherche exceptionnelles.

Renforcer la place de Paris comme acteur majeur de l'IA

M. : Parmi les startups françaises qui émergent comme Mistral AI, H ou Poolside, avez-vous une préférée ? Pour vous, laquelle est la meilleure ?

P.P. : Je me réjouis sincèrement de voir toutes ces startups émerger et prospérer à Paris, portées par des équipes brillantes. Ces talents, issus en partie des formations d'excellence françaises mais aussi de laboratoires de recherche de niveau international, s'engagent dans des aventures entrepreneuriales qui enrichissent et dynamisent l'écosystème parisien.

Etant un peu patriote, je trouve particulièrement réjouissant que ces jeunes entreprises voient le jour à Paris et en France. Cela contribue à affirmer Paris, et plus largement la France, comme un pôle majeur dans le domaine de l'intelligence artificielle, allant de la recherche académique aux startups et grands groupes. Le fait que des laboratoires de la tech américaine soient également présents à Paris, (*comme celui de Meta ndlr*), témoigne de cette dynamique. Ces laboratoires s'appuient sur une recherche académique de qualité, en tirant parti des talents, des équipes et des collaborations existantes.

Je trouve cela d'autant plus encourageant que nous évoluons dans un contexte où l'on entend parfois des propos peu élogieux sur l'Europe, la France et leur place dans le paysage international de l'IA. Pourtant, il existe ici de véritables atouts et de belles réussites. J'espère que cette dynamique se poursuivra et s'intensifiera dans les années à venir.

M.: Qu'attendez-vous du Sommet pour l'Action sur l'Intelligence artificielle ? Allez-vous participer ?

P.P. : Nous serons évidemment présents d'une manière ou d'une autre. D'après ce que je comprends, le programme est encore en cours d'élaboration, ce qui ne me permet pas d'être très précis quant à notre implication exacte. Cependant, des discussions sont en cours.

Conformément à ce que j'évoquais précédemment, nous espérons que cet événement sera un moment fort pour l'intelligence artificielle française et européenne. Il s'agit d'un sommet international, et la présence de nombreux acteurs majeurs est attendue. Le fait qu'il se tienne à Paris n'est pas anodin : c'est une opportunité pour attirer l'attention, l'intérêt, et pourquoi pas des capitaux, sur des initiatives européennes.

J'espère que les discussions ne se limiteront pas au business ou à la sécurité, mais qu'elles porteront également sur la science et les opportunités offertes par l'intelligence artificielle. C'est l'esprit que semble incarner ce sommet, et c'est essentiel pour un écosystème aussi diversifié. Il rassemble des chercheurs, des entrepreneurs, des décideurs et des investisseurs. Ce sera un moment clé pour que ces différents acteurs puissent se rencontrer et échanger, non seulement durant le sommet lui-même, mais aussi lors des événements organisés en amont, en aval et autour de celui-ci.

Il semblerait par ailleurs que des activités culturelles soient prévues le week-end. Cela me semble indispensable, car les débats sur l'intelligence artificielle ne doivent pas se limiter aux aspects techniques et scientifiques. Ce sont aussi des questions sociétales, qui touchent aux choix de société et aux modèles que nous voulons développer. Certains sujets sensibles méritent d'être abordés, comme ceux liés à la création artistique, aux contenus utilisés pour entraîner les IA ou à ceux générés

par ces dernières. Ces problématiques, riches et complexes, nécessitent une table ronde réunissant chercheurs, décideurs, juristes, artistes et autres parties prenantes.

J'espère sincèrement que ce sommet sur l'IA à Paris saura offrir cet espace de dialogue et d'échange.



À lire aussi

New Space : la startup rémoise Latitude dévoile sa feuille de route pour 2025



MADDYNEWS

La newsletter qu'il vous faut pour ne rien rater de l'actualité des startups françaises !

[JE M'INSCRIS](#)

Article écrit par Aurélie Pasquier