

AUDITIONS D'ACTEURS DU SECTEUR DE L'INTELLIGENCE ARTIFICIELLE (IA)

AUDITION #20

LA RECHERCHE SUR L'IA A PARIS

19 septembre 2024

Intervenants

- Patrick PEREZ, Directeur Général de KYUTAI

Groupes politiques :

- Emile MEUNIER, Conseiller de Paris du Groupe Les Ecologistes
- Henry MAENLER, Collaborateur du Demain, Paris !
- Jules CAPRO-PLACIDE, Collaborateur du Groupe Paris en Commun

Adjoints ou leurs cabinets :

- Pénélope KOMITES, Adjointe à la Maire de Paris chargée de l'innovation, de l'attractivité, de la prospective Paris 2030 et de la résilience – Conseillère de Paris et du 12ème arrondissement
- Kevin REVILLON, Directeur de cabinet de Pénélope KOMITES
- Alexandra MEDER, Collaboratrice de Pénélope KOMITES
- Emmanuelle LEROCHE, Conseillère au cabinet de Patrick BLOCHE
- Carla PONT, Collaboratrice de cabinet de Christophe NADJOVSKI

Administration :

- Ottavia DANINO, Cheffe de projet Innovation à la DAE

Introduction par Pénélope KOMITES, Adjointe à la Maire de Paris chargée de l'innovation, de l'attractivité, de la prospective Paris 2030 et de la résilience – Conseillère de Paris et du 12ème arrondissement

Bonjour à tous, merci Monsieur PEREZ d'être présent ce matin pour cette audition. Nous avons lancé un cycle d'auditions sur l'intelligence artificielle pour avoir une vision la plus large possible sur ce qui se passe aujourd'hui en France et à l'international, examiner les possibilités et les risques que peuvent représenter l'IA, avec l'objectif de se forger une doctrine parisienne, que la ville puisse se doter d'une stratégie IA. Nous venons de lancer une consultation en ligne auprès des Parisiens pour savoir ce qu'ils pensent de l'IA et nous aurons au mois de janvier une journée citoyenne à destination des Parisiens avec des stands, des conférences, à l'Hôtel de Ville, pour qu'ils puissent se confronter à cette nouvelle technologie.

Intervention de Patrick PEREZ, Directeur Général de KYUTAI

Merci pour cette invitation. Je vais faire une présentation générale et ensuite je m'adapterai à vos questions, à vos besoins. Je peux commencer par quelques rappels sur l'IA et vous raconter ce qu'est KYUTAI¹ et vous parler du projet qui nous a occupés dans les huit premiers mois de vie du laboratoire. L'intelligence artificielle est très médiatisée depuis quelques années alors que c'est une discipline très ancienne, qui date des débuts de l'informatique. Ce que nous appelons désormais « intelligence artificielle » est essentiellement une sous-branche de l'intelligence artificielle qui est l'apprentissage automatique, ou « *machine learning* », c'est-à-dire des outils qui consistent à faire apprendre à une machine la réalisation d'une tâche ou de plusieurs tâches, au lieu de les programmer directement, par ce que nous appelions anciennement « les systèmes experts ». C'est une méthode visant à doter les machines de certaines compétences, à travers l'utilisation quasi exclusive de données plutôt que de spécifications ou de règles. Désormais nous utilisons cette terminologie « intelligence artificielle » ou « AI », qui est commode.

Nous avons vu d'importantes révolutions ces quinze dernières années : le retour en grâce de technologies assez anciennes : « les réseaux de neurones artificiels » dans les années 2010 sont revenus sur le devant de la scène parce qu'ils marchaient très bien. Nous parlons « d'apprentissage profond ». L'adjectif « profond » ne renvoie pas à la qualité de la compréhension, mais au fait que ces réseaux de neurones sont composés d'un certain nombre de couches, et donc dotés d'une certaine profondeur. Il s'agit essentiellement d'une fonction mathématique composée d'unités de calculs inspirées de neurones biologiques. Nous les appelons « neurones artificiels » et mis ensemble, ils forment un réseau de neurones. Ce réseau peut être entraîné avant d'être utilisé et l'entraînement consiste essentiellement à fixer de façon automatique, l'intégralité des paramètres qui les définissent, qui peuvent être extrêmement nombreux, à l'aide de données, constituées dans la version initiale, d'exemple d'entrée et de sortie. C'est un programme ou une fonction, comme vous préférez. Si la tâche est de reconnaître des objets parmi une série, à la sortie nous devons retrouver les objets présents dans la série. Nous retrouvons aussi des tâches typiques de l'intelligence au sens commun, de reconnaissance ou de connaissance. Et d'autres tâches différentes, qui ne sont pas nécessairement associées à ce que nous appelons l'intelligence. Mais ce sont des techniques qui d'après un programme, permettent de réaliser un certain nombre de tâches. Dans les années 2010, les réseaux de neurones explosent. Et la deuxième révolution est, depuis deux ans pour le grand public, l'arrivée de l'IA générative, les grands modèles de langues ou les *models foundations*². Ce sont des réseaux composés de jusqu'à plusieurs centaines de milliards de paramètres fixés pendant l'entraînement pour que le réseau fasse ce qu'il a à faire. Pour l'entraîner, il doit ingurgiter un nombre considérable de données. Par ailleurs, ces très grands réseaux de neurones génèrent de la donnée : du texte ou des images. C'est « l'IA générative » ou « modèle fondation ». Ces modèles, qui sont l'objet de toute l'attention des médias et des investisseurs, de la tech en particulier, nécessitent de développer

¹ <https://kyutai.org/>

² Un modèle de fondation est un modèle d'intelligence artificielle de grande taille, entraîné sur une grande quantité de données.

énormément de moyens de calculs, de données, des expertises techniques et de l'ingénierie très pointue et recherchée. Pourquoi est-ce arrivé auprès du grand public voici deux ans ? Cela est lié au robot conversationnel Open AI, Chat GPT, qui a beaucoup surpris, y compris les gens de la communauté, par sa capacité à produire du langage parfaitement formé : orthographe, syntaxe, grammaire, permettant d'échanger de façon troublante et très humaine. Il faut garder à l'esprit que ces modèles « mécaniques » ou « computationnels » n'ont pas de réelle compréhension de ce qu'ils font, mais ils marchent. Avec ces très grands modèles arrivent... Ils sont généralistes, « *general purpose* » est la terminologie utilisée dans l'IA Act.

Ce qui est impressionnant est que l'échelle de ces modèles a permis l'émergence de modèles surprenants : des modèles qui s'entraînaient à parler ou à résoudre certaines tâches sans l'avoir acquis de façon latérale, de capacités polyvalentes. Cela les rend très puissants. Même des réseaux de neurones, Chat GPT, peuvent être utilisés pour faire de la poésie ou répondre à des questions d'ordre factuel. L'immense polyvalence de ces outils ne les rend pas très fiables. En particulier, nous pouvons nous étonner de voir que ces modèles hallucinent alors qu'à certains moments c'est ce que nous leur demandons de faire. Quand nous leur demandons d'écrire une poésie sur la ville de Paris, c'est une forme d'hallucination puisque c'est une création. Par ailleurs, une composante de ces modèles repose sur de l'aléa pour générer de la diversité. Nous pouvons demander à un robot conversationnel de régénérer sa réponse de façon différente. C'est là que nous pouvons voir qu'il hallucine, puisqu'il peut apporter deux réponses contradictoires à la même question. Pareil pour les images. Cela ouvre des perspectives considérables de productivité, de nouveaux *business* qui aiguisent l'appétit.

Cela soulève aussi des questions de dépendance à un petit nombre d'acteurs capables de les concevoir et de les modérer, avec les risques inhérents à leur très grande puissance. L'un des risques est que le contenu généré : texte, image, vidéo. Peut être dans certains cas indiscernable du contenu réel : usurpation d'identité, etc. Je passe à KYUTAI : c'est un fonds de dotation à l'heure actuelle, avec une transformation prévue, préparée en fondation reconnue d'utilité publique. C'est un fonds de dotation créé à l'initiative de trois fondateurs, Xavier NIEL, Rodolphe SAADE et Éric SCHMIDT. C'est une dotation sur fonds privés. Cela a été lancé en novembre et l'équipe initiale dont je fais partie, est une équipe de six chercheurs qui sont tous passés par des labos de la tech américaine, ce qui n'est pas anodin. Ils ont pour mission de faire de la recherche amont, plutôt généraliste, sur ces grands modèles et de partager l'intégralité des résultats. Ça s'appelle de la « science ouverte ». Partager les programmes, les modèles d'IA entraînés, partager le savoir-faire qui est derrière, en privilégiant des sujets de recherche parmi les grands défis auxquels ces modèles doivent faire face, pour qu'ils soient vraiment utiles, et de la façon la plus bénéfique possible. La fiabilité, le manque de frugalité, cela fait partie des aspects qui nous intéressent. Les plus gros modèles fonctionnent sur des machines très coûteuses, très énergivores et c'est un problème. Donc nous recherchons des modèles plus petits qui tournent sur ordinateur.

Nous sommes une petite équipe, en huit mois nous sommes passés de six à huit permanents. Nous sommes à deux pas d'ici, près de Beaubourg. Nous fonctionnons en labo de recherche. Évidemment, très vite nous nous sommes associés à une *startup*, mais nous ne sommes pas une *startup*. Nous ne faisons pas de produits, nous n'avons pas d'investisseurs, mais nous travaillons sur des objets du même type que des *startups* bien connues. Notre ambition est de devenir un acteur important de l'écosystème de l'IA à Paris, en France, en Europe et dans le monde, avec un angle singulier, celui du laboratoire à but non lucratif, mais avec des moyens considérables. Notre dotation initiale est très substantielle, ce qui nous permet d'accéder à des ressources de calcul nous permettant d'avoir de l'ambition dans nos projets de recherche. Si vous allez sur la page web KYUTAI, vous verrez un certain nombre de liens : depuis hier nous avons mis en accès libre un papier de 70 pages³ pour qui veut, avec des licences extrêmement permissives, pour faire ce qu'ils veulent avec : de la recherche, des produits, des services. Ce que fait cette IA en particulier, c'est une IA vocale avec laquelle nous pouvons parler de façon très naturelle, sans latence. C'est un prototype de recherche, une fois de plus, pas un produit. Vous pouvez même l'essayer en ligne, c'est Moshi. Depuis hier soir c'est embouteillé ! Mais vous pouvez même lui parler en anglais. C'est un modèle de langage écrit et oral en même temps, qui ne

³ <https://kyutai.org/Moshi.pdf>

repasser pas par le texte écrit pour comprendre ce que nous lui disons et pour répondre. Cela permet la rapidité, la fluidité de l'interaction, nous pouvons lui couper la parole, c'est même plutôt elle qui nous coupe la parole d'ailleurs. Cela procure une expérience très fluide, une expérience quasi humaine, avec l'expressivité qui est évidemment perdue quand nous ne repassons pas une phase de transcription textuelle. Tout ce qui est non-verbal, la voix, la prosodie, etc., sont perdus. C'est une avancée technique majeure à ce jour. Les seuls ayant ce type de modèle sont Open AI. À ceci près que leur modèle n'est pas encore déployé et le jour où il le sera, ce sera un modèle propriétaire, payant. Le nôtre est ouvert et accessible à tout le monde. C'est une grande fierté, c'est énormément de travail, surtout pour une petite équipe.

Pour vous donner un ordre d'idées, Open AI a montré, sans donner accès, une IA vocale très impressionnante également, qui commence à être testée aux États-Unis. Ils en ont parlé au mois de mai, ce qui nous a permis d'avoir quelques informations d'organisation, dessus. Ils y travaillaient depuis deux ans à 80. Nous, nous y avons travaillé six mois à huit personnes ! C'est un grand motif de satisfaction, l'équipe a été extraordinaire. Mais ce n'est qu'un début et nous allons continuer. Nous l'avons annoncé hier sur Twitter et ce matin le modèle avait été testé 300 000 fois. Nous avons mis en ligne une démo. La semaine qui a suivi, nous avons relevé plus de 200 000 connexions. Nous notons gros appétit pour ces technologies, mais pour la voix, ce n'est pas surprenant. C'est une des modalités cruciales dans l'interaction entre humains et a fortiori entre un humain et une machine ou une IA qui serait assistant, ou compagnon, suivant les cas d'usage.

Depuis juillet nous avons mis en ligne notre démo. Depuis hier nous partageons l'intégralité de la technologie. Nous sommes sollicités par de nombreux acteurs du monde des médias, de tout un tas de business. Nous leur disons : « Nous mettons les choses en *open source*, débrouillez-vous avec. » Nous pouvons collaborer pour la prise en main de ces technologies, mais ce n'est pas notre mission et ce n'est pas quelque chose pour lequel nous avons la bande passante nécessaire. C'est notre premier modèle et nous commençons déjà à travailler sur des extensions. Des extensions multilingues : nous allons le faire parler en français et nous nous intéressons à la traduction automatique à la volée à partir du même modèle. Voilà, je voulais vous présenter notre modèle et je vous tiens à votre disposition pour zoomer sur notre modèle, sur KYUTAI et plus généralement sur l'IA et sur l'écosystème parisien qui est dans une dynamique absolument magnifique. Paris peut s'enorgueillir d'avoir en Europe et dans le monde, l'un des écosystèmes d'IA les plus stimulants.

Pénélope KOMITES, Adjointe à la Maire de Paris chargée de l'innovation, de l'attractivité, de la prospective Paris 2030 et de la résilience – Conseillère de Paris et du 12ème arrondissement

Merci beaucoup, je vais ouvrir aux questions.

Emile MEUNIER, Conseiller de Paris du Groupe Les Ecologistes

Merci pour cette présentation et pour le travail que vous menez. Vous répondez à l'une de mes inquiétudes principales : comment garder notre indépendance en tant que pays ou que continent par rapport aux géants chinois, américains et autres ? J'ai l'impression que nous sommes déjà une colonie numérique et que cela pourrait s'aggraver encore plus. Et donc une des réponses, c'est ce que vous faites. Mais pardonnez-moi, vous êtes huit parmi les plus talentueux au monde, félicitations ! Mais vous n'êtes que huit, face à des armées aux États-Unis ! Comment rester dans la course, comment ne pas se faire déclasser ? Et deuxième point : comment la ville de Paris peut-elle se positionner par rapport à cela ? Ce peut être en aidant l'écosystème d'une manière ou d'une autre, même si je pense que c'est le rôle de l'État avant tout, à mon avis. Mais nous en tant que service public, comment voyez-vous nos relations avec l'intelligence artificielle ?

Intervention de Patrick PEREZ, Directeur Général de KYUTAI

La place de l'Europe dans la course : l'IA n'est que le dernier développement en date de la plateformes et de la numérisation du monde. Certains combats sont difficiles désormais à gagner, effectivement. Les géants d'internet sont aux États-Unis et en Chine. Nous aurons du mal à

ratrapper ces géants-là. Et par ailleurs ce sont parmi les mieux placés au monde pour capitaliser et contribuer à des avancées en IA. Parce qu'ils ont des moyens illimités, ils ont accès à des données illimitées, ils savent faire des choses à grande échelle et ils ont réussi à constituer parmi les équipes de recherche et d'ingénieurs les meilleures au monde, grâce à tous les atouts que j'ai mentionnés. L'argent, les machines et la donnée. Dit comme ça, ce pourrait être décourageant. En même temps, ce n'est pas nécessairement dans les plus grandes équipes... En même temps la bataille génère un certain encombrement. Dans les 18 mois écoulés, il est intéressant de voir le fait qu'Open AI qui était au départ essentiellement un laboratoire à but non lucratif, une petite pousse sortie de nulle part a réussi à ébranler Google en quelques années. En capitalisant sur des technologies en partie inventées par Google ! Donc cette histoire de grands groupes monopolistiques et archi-dominants qui se font dépasser par de petits acteurs sortis de nulle part continue à arriver. Et Open AI est impressionnant de ce point de vue. Je prends un autre exemple : je viens de passer six ans à faire de l'IA dans le monde automobile chez Valéo.

Chez Valéo, les acteurs historiques regardaient en rigolant Tesla. Or, Tesla a révolutionné le monde automobile avec une approche *startup* et natif digital. La question s'était déjà posée avec les plateformes et effectivement nous avons perdu la bataille des plateformes et des réseaux sociaux. Nous devons tirer les leçons de ce qui s'est passé il y a 20 ans avec les plateformes et les réseaux sociaux. Cela doit s'organiser à tous les échelons : l'État, l'Europe. Les choses sont compliquées à l'heure actuelle en Europe autour de la réglementation, mais déjà, la volonté politique de l'Europe de faire bloc sur l'IA avec des investissements... Je fais allusion au rapport de Mario DRAGHI⁴ sur notre perte de vitesse. Donc des questions d'investissements, de volonté politique et des atouts. Les atouts ce sont les écosystèmes : Paris, Londres, Cambridge, la Suisse avec Lausanne et Zurich, l'Allemagne avec des *hubs* très forts, les Pays-Bas. Nous avons des atouts avec des systèmes académiques et privés. D'ailleurs les GAFAM ne s'y sont pas trompés : ils ont dans toutes ces villes-là des labos de recherche. Et pas que de façon prédatrice, d'ailleurs : ils contribuent à l'écosystème quand même, parce qu'ils savent qu'ils ne peuvent pas arriver et vider les laboratoires publics de leur substance. Ils ont besoin d'une synergie. Ce qui s'est passé à Paris en particulier, une grande partie de mon équipe vient du laboratoire de Facebook, qui a Paris a bénéficié de l'écosystème qui était déjà là, de très bons laboratoires publics et privés qui existaient déjà. Et ils ont rendu à cet écosystème, en formant des gens, des gens qui sont pour un certain nombre, restés et d'autres qui sont allés créer des *startups* : Mistral, ou des laboratoires comme le mien. Des gens ont envie d'être dans ces aventures nouvelles qui sortent de terre. L'une des motivations est patriote, souveraine, en disant que cela bénéficie aussi à l'Europe et pas qu'aux géants de la tech américaine. Et un autre point important est que la course folle sur ces sujets entre les géants de la tech américaine, fait qu'une partie de leurs laboratoires de recherche se sont en partie refermés. Ils ont commencé à moins publier, à moins partager les codes, ce qui fait que pour un certain nombre de chercheurs, ce n'est pas acceptable.

Une chose n'est pas négociable : la libre diffusion des produits. Le fait que dans mon équipe ce ne soit que des gens qui sont passés par la tech américaine, fait que certains disent « Je ne veux pas être dans un laboratoire avec 3 000 chercheurs, énormément de politique, de la fébrilité parce que c'est la compétition avec Facebook, Microsoft et Open AI et nous commençons à être restreints dans notre liberté de publier ». C'est ainsi que des personnes très talentueuses veulent créer de nouveaux laboratoires, créer des boîtes, retourner dans la recherche publique pour ceux qui en viennent. Dans l'équipe initiale de six, nous sommes deux anciens chercheurs de labo. J'avais tellement d'années que j'ai atteint les limites de disponibilités, mon collègue avait encore de la disponibilité, mais nous avons des possibilités de retourner dans la recherche publique. Cela se fait, j'ai fait trois fois cela dans ma carrière, d'aller vers le public puis vers le privé et c'est très extrêmement intéressant. Et pour revenir à Paris, je n'ai pas de suggestions très concrètes, mais déjà, ce que nous faisons aujourd'hui, se voir et discuter, est très important. Cet écosystème de l'IA à Paris est très dynamique, extrêmement intéressant, mais encore fragile. Quand les investissements nécessaires sont élevés nous avons besoin de fonds d'investissements étrangers, côté États-Unis ou Moyen-Orient. Ce n'est pas un problème en soi, mais la difficulté va être ensuite de garder les *startups* sur Paris. Ensuite, je prends l'exemple du calcul : les machines que requiert le déploiement de grands modèles coûtent énormément d'argent,

⁴ https://commission.europa.eu/document/97e481fd-2dc3-412d-be4c-f152a8232961_en

mais même avec de l'argent, peu d'entreprises peuvent fournir ces machines. Nous travaillons, peut-être sans surprise, avec Scaleway du groupe Iliad, qui ont leurs machines en Région parisienne. Même cette question d'où mettre les *datacenters* pose des problèmes techniques d'approvisionnement, d'électricité. Ce que j'entends des *providers* de Scaleway est que monter des *datacenters* en France a l'air d'être un parcours du combattant. Je ne veux pas dire qu'il faudrait en faire pousser à tous les coins de rue, cela soulève une vraie question qui est celle du coût écologique. Mais pour ce qui est de la souveraineté et d'ancrage de l'écosystème, trois choses cruciales sont : d'avoir accès aux ressources de calcul suffisantes, avoir les talents et les garder, qu'ils ne soient pas tous dans les boîtes américaines ou chinoises. Et concernant les données, il s'agit d'un sujet vraiment important. Les données servant à l'entraînement de grands modèles de langues déterminent ce que l'IA comprend du monde. Et ces viennent pour l'essentiel d'internet.

Mettons de côté les questions de copyright que cela pose. Cela est très biaisé, c'est la langue anglaise, voire nord-américaine. Là, nous avons un espoir. Ces données se font à très grande échelle, elles servent au pré-entraînement des modèles. Mais ensuite, beaucoup de choses se font avec des données en beaucoup plus petite quantité, mais beaucoup plus grande qualité. L'accès à ces données-là coûte beaucoup d'argent, soit elles existent déjà, mais ne sont pas accessibles. Depuis six mois je discute avec des acteurs institutionnels et privés du monde des médias. Demain je rencontre l'INA avec des données immenses et ce sont des données dont ils sont dépositaires, pas propriétaires. Il existe de nombreuses questions autour de l'accès aux données. Les groupes de médias publics comme Radio France ou France Télévision ou privés... Les deux fondateurs KYUTAI ont chacun un groupe média donc nous discutons avec eux. Mais je constate que ce sont des personnes qui ont été échaudées, elles se sont fait tailler les croupières depuis 20 ans par les plateformes donc ils ne vont pas se faire avoir une seconde fois. Donc ils sont un peu frileux. Mon espoir avec ma casquette KYUTAI de laboratoire non lucratif est d'être reçu différemment. Mais ils se disent que tout le monde va profiter de *l'open source*, y compris la Chine et les États-Unis. Oui, mais c'est ça, *l'open source* ! Il s'agit d'un sujet compliqué donc il faut y aller soigneusement, mais les institutions peuvent avoir des choses à discuter avec les détenteurs de données de qualité.

Emile MEUNIER, Conseiller de Paris du Groupe Les Ecologistes

Hier nous avons auditionné l'APUR, notre agence parisienne d'urbanisme, qui depuis 1970 a eu le réflexe de numériser leurs données. Donc ils ont des données d'une extrême richesse et maintenant avec l'IA ils commencent à produire de la valeur, de nouveaux champs de recherche, de la cartographie...

Intervention de Patrick PEREZ, Directeur Général de KYUTAI

Effectivement ce sont des données qui valent beaucoup et encore sous-estimée. Lorsque des groupes, des acteurs en Espagne, aux États-Unis, etc., signent des contrats, non exclusifs heureusement avec Open AI pour avoir accès à leurs données. Ils sous-estiment parfois la richesse de ce qu'ils ont, eux. Et ce qui serait dommage serait que des acteurs européens ne bénéficient pas de l'accès à ces données-là. Je pense qu'à tous les niveaux : institutionnels, politiques et privés, travailler cette question du partage de la donnée est intéressant.

Pénélope KOMITES, Adjointe à la Maire de Paris chargée de l'innovation, de l'attractivité, de la prospective Paris 2030 et de la résilience – Conseillère de Paris et du 12ème arrondissement

Merci, j'avais quelques questions : pourquoi avoir lancé Moshi en anglais ? Après, nous avons bien vu au fur et à mesure des auditions que la question des données est essentielle. Et nous partageons le fait qu'il faille un accès aux données. Comment verriez-vous le rôle de la ville, dans ce partage des données ? La ville a mis ses données en *open data* depuis un certain temps, mais devrions-nous signer des partenariats avec des acteurs qui gravitent autour de la sphère publique pour travailler ensemble, pour une mise en partage des données, systématiques ? Quelle peut être notre influence sur cet accès aux données ?

Intervention de Patrick PEREZ, Directeur Général de KYUTAI

C'est une question difficile : je connais les intrications politiques et institutionnelles sur des acteurs basés à Paris, des acteurs nationaux comme l'INA ou la BNF. Mais peut-être que la ville de Paris peut être un médiateur ou un facilitateur. Après, le fait que Paris brasse toute une série de données... Les données sont sensibles, elles renvoient à des personnes, des individus ou des organisations. Il faut faire attention, mais j'imagine que la ville de Paris a accès à des données hétérogènes, qui mises ensemble, peuvent être intéressantes. Je suis intéressé aussi de voir quels sont les cas d'usage de ces acteurs-là. Par exemple un groupe de média a ses clients, consommateurs ou lecteurs. Mais eux-mêmes ont aussi des besoins d'outils spécifiques pour eux, en tant que professionnels. Cela est vrai dans tous les métiers. Lorsque nous discutons avec des gens qui s'intéressent à l'IA pour le juridique, les ressources humaines, etc. Ce sont des problématiques métiers, quand même. J'imagine qu'à la ville de Paris vous ne manquez pas de questions de ce type-là, peut-être avec des questions assez spécifiques. L'échelle est plus importante que dans d'autres métropoles. Je discute avec des groupes de médias, je leur dis que nous pouvons aussi avoir des groupes de modèles avec en contrepartie, le fait que certaines versions du modèle soient pour vous, pour vos besoins en interne. Cela nécessite de bien comprendre les cas d'usage et ce que nos technologies, qui sont assez généralistes, peuvent ou pas, pour ces cas d'usage.

Pénélope KOMITES, Adjointe à la Maire de Paris chargée de l'innovation, de l'attractivité, de la prospective Paris 2030 et de la résilience – Conseillère de Paris et du 12ème arrondissement

Quand vous dites que l'écosystème parisien est en ébullition et en même temps fragile, comment verriez-vous le rôle de la ville pour soutenir cet écosystème ?

Intervention de Patrick PEREZ, Directeur Général de KYUTAI

Des lieux peut-être, qui reposent sur des initiatives privées, je pense par exemple à Station F où l'annonce de Kyutai avait été faite. Ce n'est pas anodin que nous soyons à deux pas d'ici et c'est vrai pour un Google, un Mistral, etc. Il est essentiel d'être très central pour l'attractivité.

Emile MEUNIER, Conseiller de Paris du Groupe Les Ecologistes

Et moi, Porte de la Chapelle, nous avons des dizaines de milliers de mètres carrés et personne n'en veut !

Pénélope KOMITES, Adjointe à la Maire de Paris chargée de l'innovation, de l'attractivité, de la prospective Paris 2030 et de la résilience – Conseillère de Paris et du 12ème arrondissement

Et Mistral ne veut pas quitter le centre de Paris bien qu'ils commencent à être à l'étroit.

Intervention de Patrick PEREZ, Directeur Général de KYUTAI

Ce serait contre-productif ! Les gens dont nous parlons, ce sont un peu des divas. Cela peut être agaçant, mais ces ingénieurs sont courtisés de façon dingue et le moindre élément a son importance. Le fait d'être au centre de Paris n'est pas négociable. Faciliter cela, de façon raisonnable, peut être intéressant. Et l'écosystème est aussi académique, avec la recherche publique. Dans notre équipe nous avons des doctorants, inscrits dans des écoles doctorales, encadrés par des directeurs de thèse dans des labos publics en particulier dans Paris. À Paris, nous avons au moins deux *clusters* ; ce sont des choses très importantes. Ce sont même eux qui ont plus besoin d'aide, que nous. Côté Sorbonne, ils ont un lieu, très chouette d'ailleurs. Prairie j'ai l'impression que ça a été plus compliqué de trouver un lieu. En tout cas, choyer la partie académique est important. Et aussi, l'accès aux machines. C'est lié à la stratégie nationale. GENCI⁵ est l'agence qui opère le supercalculateur Jean ZAY entre autres et qui est indispensable pour que l'écosystème en particulier académique, français, ait accès à des ressources de calcul décentes pour faire de l'IA. Ces machines existent, peut-être des synergies sont-

⁵ Grand équipement national de calcul intensif : <https://www.genci.fr/>

elles à regarder de ce côté-là, en particulier tournées vers la partie publique de l'écosystème parisien. Nous avons de très très beaux labos d'IA à Paris. Certains sont assez loin, ça reste difficile d'aller les vendre, de mon point de vue.

Pénélope KOMITES, Adjointe à la Maire de Paris chargée de l'innovation, de l'attractivité, de la prospective Paris 2030 et de la résilience – Conseillère de Paris et du 12ème arrondissement

Et Moshi ?

Intervention de Patrick PEREZ, Directeur Général de KYUTAI

Ah oui, pourquoi en anglais ? Parce que la partie initiale repose sur de l'entraînement de texte et d'audio sur internet. Et par ailleurs, pour pouvoir se comparer, dans le papier que nous avons sorti hier et qui fait 70 pages, nous avons énormément de tableaux. Ce sont des benchmarks quasi académiques donc c'est en anglais. Nous commençons à faire celui en français, mais pour cela il nous faut de la donnée audio en français, qui est moins sur internet. C'est pour cela que nous discutons depuis un moment avec des groupes français, espagnols : pour cette raison. Mais voilà pourquoi nous avons fait Moshi en anglais : c'est plutôt l'anglais qui est accessible en termes de données et la partie benchmarking pour la comparaison. Nous avons beaucoup de demandes pour que l'outil soit en français, en arabe, dans des langues du continent indien. Ce sera sans doute possible, mais pas avec nous. On va donner les recettes pour que les gens adaptent le modèle.

D'abord, les sommes engagées sont sans récurrence. CMA-CGM, ce sont 100 millions d'euros et ILIAD ce sont 100 millions d'euros, ce sont des sommes considérables, mais il n'est pas prévu qu'ils en remettent ! Donc nous avons un certain nombre d'années devant nous avec la dotation initiale. Si nous élargissons beaucoup la puissance de calcul, cela va réduire la durée. Donc une chose que je commence à faire, c'est prospecter pour d'autres donations. Par ailleurs, le fait que la valeur soit entièrement distribuée, n'empêche pas de créer des *business*, que des choses naissent autour de Kyutai qui ne soient pas Kyutai, qui permettent de ramener de la valeur aux fondateurs, mais aussi de pérenniser Kyutai sous forme de rentrées d'argent. Même si en étant une fondation, cela limite la dimension commerciale qui peut en être faite. Et nous n'avons pas d'urgence sur ce point. Ce qui est important c'est que les équipes restent concentrées sur ces recherches et ne soient pas distraites par des recherches de financement, faire du *business*, ce n'est pas l'idée.

Jules CAPRO-PLACIDE, Collaborateur du Groupe Paris en Commun

Qui a-t-il à gagner à cela ? Et comment faire pour en quelques mois arriver à ce niveau, outre le talent ? Et sur les ressources humaines, vous êtes six au départ, formés aux USA, donc est-ce que la formation en France est nulle ?

Intervention de Patrick PEREZ, Directeur Général de KYUTAI

Nous n'avons pas été formés aux USA. Nous avons été formés par les labos de la tech américaine, mais pas aux USA. Nous sommes tous les six sortis de trois écoles françaises. Et nous avons tous travaillé pour des labos de la tech américaine. Et à part un, c'était tout en France, sauf un en Angleterre. Certains n'ont même pas quitté la France. Donc la formation en France est tout sauf nulle, elle est exceptionnelle ! A Paris, le Master MBA est le meilleur Master de recherche en IA, qui est victime de son succès parce qu'ils sont à plus de 200 ou 300 étudiants, tous les polytechniciens qui veulent y aller ne peuvent pas tous y aller parce qu'il n'y a plus de place. La formation est impeccable. Et est-ce que nous avons recruté ? Oui. Nous recrutons peu parce que nous sommes hyper exigeants. Et par ailleurs nous ne sommes pas les seuls à leur courir après. Et nous avons en face de nous des gens qui ont quelque chose que nous n'avons pas, qui est le recours aux actions gratuites ou aux parts dans une société, parce que nous, nous sommes *non-profit*. Quelqu'un qui sort de thèse et qui va chez Google ou Facebook, les packages sont monstrueux. Nous n'avons pas cet outil-là, mais nous avons d'autres atouts. Donc oui, nous cherchons à recruter, mais pas à n'importe quel prix et nous n'avons pas hâte. Il est normal que les *startups* se lancent dans des courses folles pour faire vivre leurs produits. Faire un produit c'est autre chose que faire un prototype de recherche. Donc nous, nous n'avons pas cette problématique. Et votre question, sur pourquoi faire cela : nous avons donc trois fondateurs

donateurs. Le chiffre officiel est près de 300 millions pour la totalité. En fait c'est moins que cela. Et pourquoi ces gens-là font cela ? Je pense qu'ils le font pour des raisons qui leur sont propres et différentes : dans le cas d'Éric Schmitt, il voit que tout Américain qu'il est, pour la bonne santé de l'écosystème il faut que l'IA soit également présente en Europe pour des raisons géopolitiques. Il a vraiment une vision géopolitique de l'IA qui l'a encouragé à faire cela. Et les deux autres, deux Français, ont le filtre français, mais en faisant cela ils contribuent à un écosystème dans lequel ils ont par ailleurs d'autres activités, d'autres investissements, donc c'est un cercle vertueux. Et cela leur permet d'avoir un accès direct à de la connaissance, non pas un accès privilégié puisque nous partageons avec tout le monde, mais tout de même, nous échangeons avec eux, ce qui est la moindre des choses vu les investissements.

Jules CAPRO-PLACIDE, Collaborateur du Groupe Paris en Commun
Comment avez-vous fait pour, en quelques mois, sortir Moshi ?

Intervention de Patrick PEREZ, Directeur Général de KYUTAI

D'abord c'est énormément de travail et peu de distractions. Mais parfois nous avons beaucoup ri avec Moshi dans l'Open Space. Mais c'est un concentré de personnes qui ont un réel savoir-faire. Nous avons parmi les meilleurs experts audios de la planète, un de chez Google et un de chez Facebook, jusqu'à présent ils étaient concurrents et maintenant ils sont ensemble. Nous avons un des meilleurs experts de langage de la planète, l'un de ceux qui ont fait Lama. Ces gens savaient déjà faire une partie de ce que nous avons fait. Ce que nous ne savions pas faire, c'était mettre ensemble l'audio et le langage. C'était un pari. C'est ce qui est intéressant avec la recherche c'est que nous prenons plus de risques que dans une *startup*. Les *startups* qui faisaient des LLM en ce moment savaient déjà le faire. Ce n'est pas sous-estimer le challenge : en six mois nous avons réuni l'expertise et l'expérience antérieure, mais la part d'ingénierie de faire tourner des modèles, filtrer les données, les récupérer... C'est une ingénierie de malade ! Et là en n'étant que six... Moi je ne programme plus, j'étais parmi les six, celui qui ne programmait pas. Ce n'était pas gagné ! C'est assez risqué, certaines boîtes prennent des risques. S'ils n'ont pas déjà la quasi-certitude qu'ils vont faire quelque chose qu'ils savent faire, ce peut être assez risqué vu les montants impliqués.

Emile MEUNIER, Conseiller de Paris du Groupe Les Ecologistes

Toujours sur cette particularité que nous avons en France d'avoir un bel écosystème grâce à nos écoles de mathématiques comme Normale Sup et Polytechnique... De l'autre côté je lis beaucoup que notre niveau de mathématiques se dégrade. Est-ce que vous percevez qu'il n'y a pas d'inquiétudes à avoir, sur le niveau en mathématiques ?

Intervention de Patrick PEREZ, Directeur Général de KYUTAI

Je vais répondre en tant que citoyen français, père de famille. J'ai des enfants qui ont fait des études scientifiques. Si, le niveau en maths baisse et de façon flagrante dans le secondaire et la chute s'accélère. *A fortiori* pour les filles : c'est une catastrophe ! Donc oui, cela va mettre un peu de temps avant que cela ne se voie dans les niveaux supérieurs. Mais au niveau lycée, oui. Mais la partie scientifique des prépas reste archi solide. Ma fille est en deuxième année à Polytechnique et avec qui j'ai passé mon temps à faire des maths quand elle était en Maths sup. Le niveau reste stratosphérique et l'exigence est là. Mais l'écart se creuse entre cette formation d'élite et le reste. Et l'écart entre filles et garçons n'est pas acceptable.

Ottavia DANINO, Cheffe de projet Innovation à la DAE

Je crois beaucoup au fait que notre atout est l'écosystème et la synergie entre tous les acteurs, à Paris notamment. Je crois d'ailleurs qu'Open AI lance un bureau à Paris. Et vous avez dit que l'ambition de Kyutai est d'être un acteur important de cet écosystème : avec qui travaillez-vous en partenariat, outre Scaleway ?

Intervention de Patrick PEREZ, Directeur Général de KYUTAI

Nous ne travaillons avec personne. Même avec Scaleway, ce n'est pas un partenariat. C'est crucial d'avoir eu accès à leurs machines dès le premier jour, mais ce n'est pas un partenariat en termes de développement. C'est un fournisseur crucial. Des partenariats plus stratégiques : nous avons eu énormément de sollicitations et nous les avons déclinées ou mises en *stand-by*, essentiellement pour des raisons de disponibilités. Pas de distraction. Ça ne veut pas dire qu'il n'y en aura pas. Nous en aurons très rapidement avec des labos publics : chaque thèse que nous allons lancer sera une collaboration scientifique avec un labo public. Nous avons des collaborations avec l'IRCAD, l'université, les Ponts et Chaussées. À chaque fois ce sont des collaborations avec un chercheur ou une chercheuse, un directeur, des doctorants entre les deux endroits qui font le lien et un projet développé ensemble, mais nous n'avons pas de partenariat public-privé en revanche.

Clôture de la réunion par Pénélope KOMITES, Adjointe à la Maire de Paris chargée de l'innovation, de l'attractivité, de la prospective Paris 2030 et de la résilience – Conseillère de Paris et du 12ème arrondissement

Je vous remercie beaucoup, c'était passionnant. A bientôt pour d'autres échanges.

[Fin de l'audition à 12 h 13]