



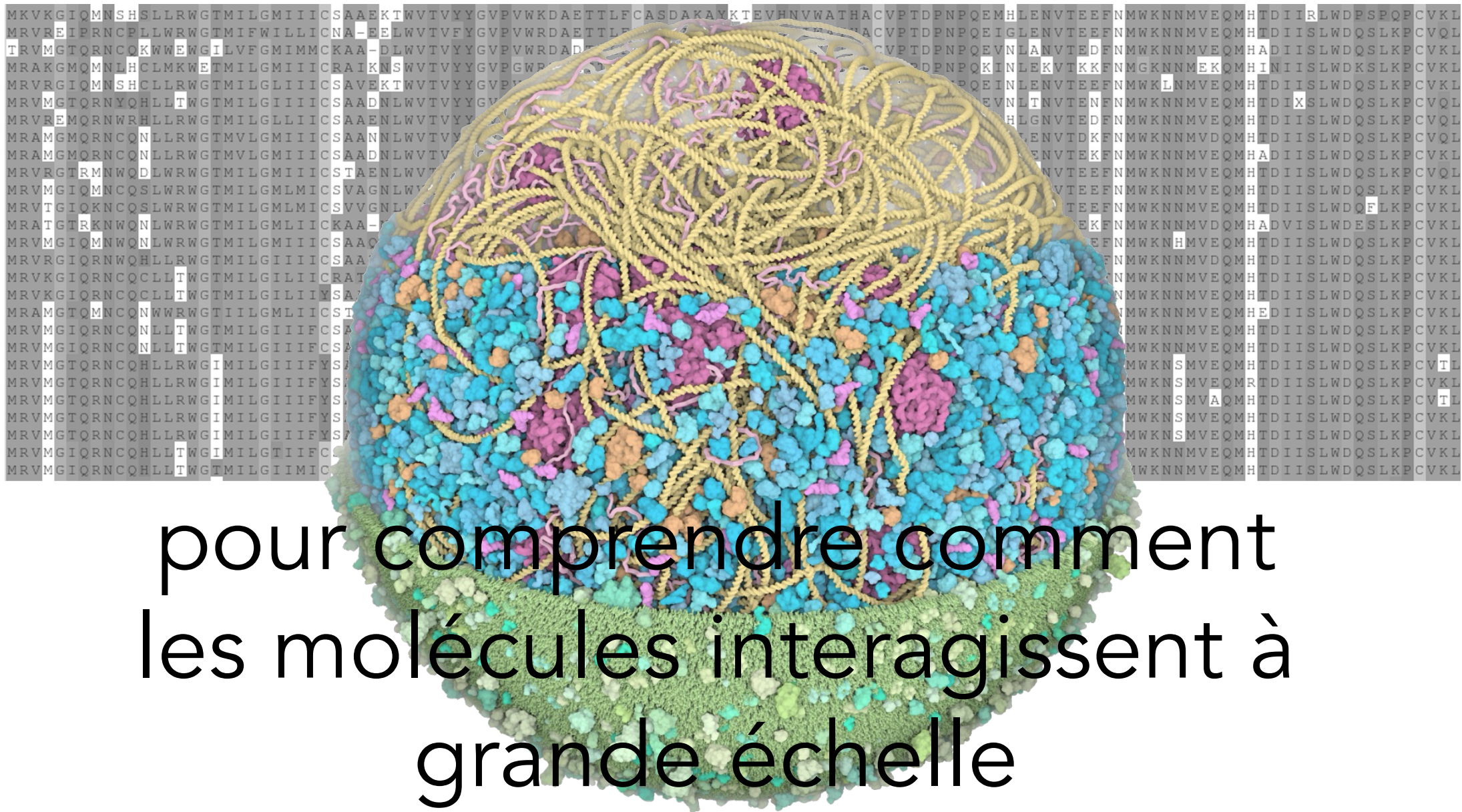
Alessandra Carbone

Professeur d'Informatique, Sorbonne Université

Directrice du Laboratoire de Biologie Computationnelle, Quantitative et Synthétique,
Sorbonne Université et CNRS

MKVKG IQMNSHSLLRWGT MILGMI IICSA AEK TWVTVYYGV PVWKDAETTLFCASDAKAYKTEVHNW WATHACVPTDPNPQEMHLENVTEEFNMWKNNMVEQMHTDIIRLWDPSPQPCVKL
MRVREIPRNCPLLRWGT MIFWILLICNA-EELWVTVFYGV PVWRDAETTLFCASDAKAYDTEVHNW WATHACVPTDPNPQEIGLENVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCVQL
TRVMGTQRNCQKWWEWGILVFGMIMMCKAA-DLWVTVYYGV PVWRDAETTLFCASDAKAYATEKHNVW WATHACVPTDPNPQEVNLANVTEDEFNMWKNNMVEQMHTDIISLWDQSLKPCVKL
MRAKGMQMNHLHCLMKWETMILGMI IICRAIKNSWVTVYYGV PGWRDAKTTFFWASNAKAYKSKGHKQVWATLACVPPDPNPQKINLEKVTKKFNMGKNNMEKQMHINIISLWDKSLKPCVKL
MRVRGIQMNSHCLLRWGT MILGLI IICSAVEKTWVTVYYGV PVWKDAETTLFCASDAKAYTTEMHNW WATHACVPTDPNPQEIINLENVTEEFNMWKLNMVEQMHTDIISLWDQSLKPCVQL
MRVMGTQRNYQHLLTWGT MILGI I IICSAADNLWVTVYYGV PVWKDAETTLFCASDAKAYDKEVHNW WATHACVPTDPNPQEVNLTINVTEFNMWKNNMVEQMHTDIISLWDQSLKPCVQL
MRVREMQRNWRHLLRWGT MILGLLI IICSA AENLWVTVYYGV PVWKDAETTLFCASDAKAYVTEKHNVW WATHACVPTDPNPQEIHLGNVTEDEFNMWKNNMVEQMHTDIISLWDQSLKPCVQL
MRAMGMQRNCQNLLRWGT MVLGMI I IICSAANNLWVTVYYGV PVWKDAETTLFCASDAKAYDTEVHNW WATHACVPTDPNPQEMKLENVTDKFNMWKNNMVDQMHTDIISLWDQSLKPCVQL
MRAMGMQRNCQNLLRWGT MVLGMI I IICSAADNLWVTVYYGV PVWKDAETTLFCASDAKAYDTEVHNW WATHACVPTDPNPQEMKLENVTEKFNMWKNNMVEQMHTDIISLWDQSLKPCVKL
MRVRGTRMNWQDLWRWGT MILGMI I IICSTAENLWVTVYYGV PVWKDAETTLFCASDAKAYDTEMHNW WATHACVPTDPNPQEIINLKNVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCVQL
MRVMGIQMNCQSLRWGT MILGMLMI CSVAGNLWVTVYYGV PVWKEADTTLFCASNARAYDTEVHNW WATHACVPTDPNPQEIIDLNVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCVKL
MRVTGIQKNCQSLRWGT MILGMLMI CSVVGNLWVTVYYGV PVWKEADTTLFCASNARAYDTEVHNW WATHACVPTDPNPQEIIDLNVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCVKL
MRATGTRKNWQNLWRWGT MILGMLI ICKAA-NLWVTIYYGV PVWKDAETTLFCASDAKAYEKEVHNW WATHACVPTDPNPQEIINLGNVTEKFNMWKNNMVDQMHTDIISLWDES LKPCVKL
MRVMGIQMNWQNLWRWGT MILGMI I IICSA AQLWVTVYYGV PVWKEADTTLFCASDAKAYNTEMHNW WATHACVPTDPNPQEIINLENVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCVKL
MRVRGIQRNWQHLLRWGT MILGI I I IICSA AENMWVTVYYGV PVWRDAETTLFCASDAKGYETEKHNVW WATHACVPTDPNPQEIHLNVTEEFNMWKNNMVDQMHTDIISLWDQSLKPCVKL
MRVKG IQRNCQCLLTWGT MILGILI IICRATENLWVTVYYGV PVWKDAKTTLFCASDAKAYETEKHNVW WATHACVPTDPNPQEIHLNVTEDEFNMWKNNMVEQMHTDIISLWDQSLKPCVKL
MRVKG IQRNCQCLLTWGT MILGILI I ICSA AEDLWVTVYYGV PVWKDAKTTLFCASDAKAYETEKHNVW WATHACVPTDPNPQEIHLNVTEDEFNMWKNNMVEQMHTDIISLWDQSLKPCVKL
MRAMGTQMNCQNWWRWGT IILGMLI IICSTAEDLWVTVYYGV PVWRDAETTLFCASDAKAYDTEMHNW WATHACVPTDPNPQEIINLGNVTEKFNMWKNNMVEQMHTDIISLWDQSLKPCVKL
MRVMGIQRNCQNLLTWGT MILGI I I IICSATDNLWVTVYYGV PVWKDAETTLFCASDAKAYATEKHNVW WATHACVPTDPNPQEIHLNVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCVKL
MRVMGIQRNCQNLLTWGT MILGI I I IICSATDNLWVTVYYGV PVWKDAETTLFCASDAKAYATEKHNVW WATHACVPTDPNPQEIHLNVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCVKL
MRVMGTQRNCQHLLRWGT IMILGI I IIFYSA AENMWVTVYYGV PVWKDAETTLFCASDAKAYETEKHNVW WATHACVPTDPNPQEIHLNVTEQFNMWKNSMVEQMHTDIISLWDQSLKPCVTL
MRVMGTQRNCQHLLRWGT IMILGI I IIFYSA AENMWVTVYYGV PVWRDAETTLFCASDAKAYETERHNW WATHACVPTDPNPQEIHLNVTEQFNMWKNSMVEQMHTDIISLWDQSLKPCVKL
MRVMGTQRNCQHLLRWGT IMILGI I IIFYSA AENMWVTVYYGV PVWKDAETTLFCASDAKAYETEKHNVW WATHACVPTDPNPQEIHLNVTEQFNMWKNSMVAQMHTDIISLWDQSLKPCVTL
MRVMGTQRNCQHLLRWGT IMILGI I IIFYSA AENMWVTVYYGV PVWKDAETTLFCASDAKAYETERHNW WATHACVPTDPNPQEIHLKNVTEQFNMWKNSMVEQMHTDIISLWDQSLKPCVKL
MRVMGTQRNCQHLLRWGT IMILGI I IIFYSA AENMWVTVYYGV PVWRDAETTLFCASDAKAYETERHNW WATHACVPTDPNPQEIHLKNVTEQFNMWKNSMVEQMHTDIISLWDQSLKPCVKL
MRVMGTQRNCQHLLTWGT IMILGI I I IICSAVENLWVTVYYGV PVWRDAETTLFCASDAKAYETEKHNVW WATHACVPTDPNPQEIHLDNVTEKFNMWKNNMVEQMHTDIISLWDQSLKPCVKL
MRVMGIQRNCQHLLTWGT IMILGI I I IICSA AENLWVTVYYGV PVWKDAETTLFCASDAKAYETEKHNVW WATHACVPTDPNPQEIHLNVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCVKL

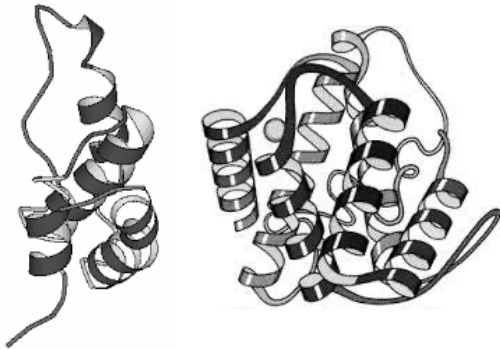
Décoder les signaux biologiques
à partir des séquences



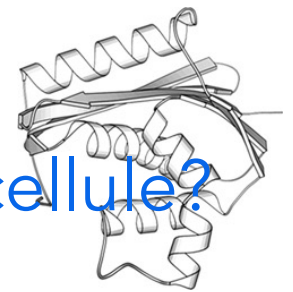
Quelles protéines interagissent entre elles dans la cellule?

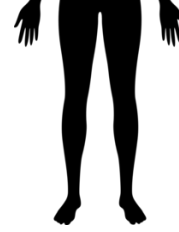
2006-2013

modélisation moléculaire



Quelles protéines interagissent entre elles dans la cellule?



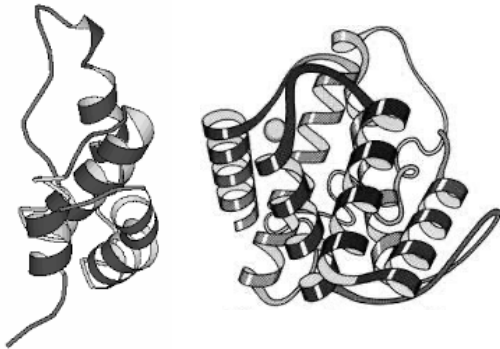


2006-2013

2023

modélisation moléculaire

apprentissage profond



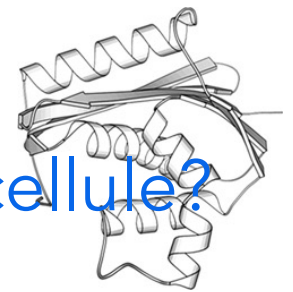
MTLRKLLTGELLTLASR
QQLIDWMEADKVG
PLLSALPAGWFIADK
SGAGERGSRGIPEDRN
RGLAASCWFIADKCF
VLALLTLAKLEKDFRLL
GLC...



MRQQLIDWGELLTLAS
ADKVGPLLSALPLE
KDFRLLGAGWFIADKS
GANDTLRKLLTGERGS
SCWFERKCLAKLCFC
VLALLRGIPEDRNRGL
AAT...



Quelles protéines interagissent entre elles dans la cellule?



2006-2013

modélisation moléculaire

3 ans

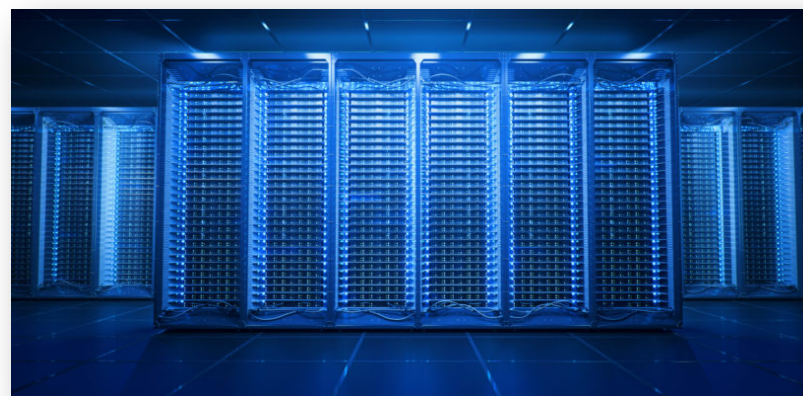


avec World Community Grid (IBM)

2023

apprentissage profond

1 semaine



avec un cluster de GPU à SU

2006-2013

modélisation moléculaire

3 ans



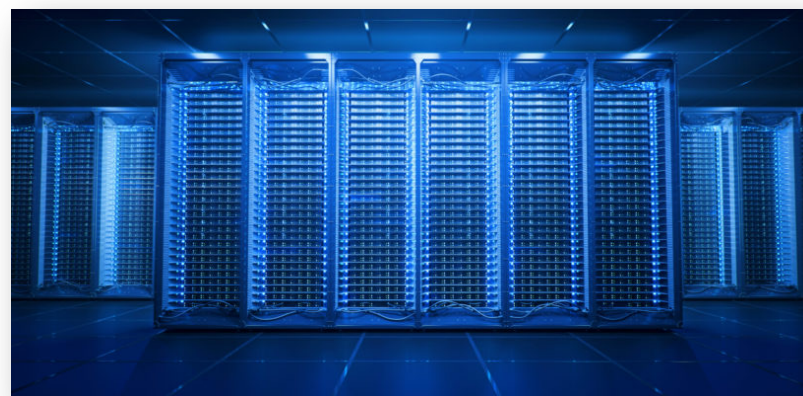
avec World Community Grid (IBM)

2 000 x 2 000

2023

apprentissage profond

1 semaine



avec un cluster de GPU à SU

20 000 x 20 000

2006-2013

modélisation moléculaire

Modèle

abstrait de la réalité, conçu pour
démontrer, expliquer et prédire

2023

apprentissage profond

Données

décrivant une vérité partielle,
qui nous permettent
de découvrir des patterns inattendus,
avec une efficacité surprenante,
et que l'on peut utiliser pour prédire !

L'IA a apporté un changement fondamental de paradigme

Q: faudra-t-il conserver ces deux faces du raisonnement – la modélisation théorique et l'apprentissage par les données – dans le futur? Quel poids leurs donner?

On se priverait de connaissance cruciale si on ne développait pas les deux, l'une à côté de l'autre !

1. L'IA nous fait changer les données que l'on manipule

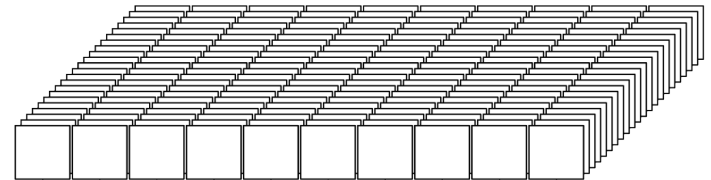
Séquences

N	E	V	F	H	T	G	I	K	Q
---	---	---	---	---	---	---	---	---	---

Objets réels

Embeddings

Encodages numériques



"Objet augmenté"
doté d'"expertise"

1. L'IA nous fait changer les données que l'on manipule

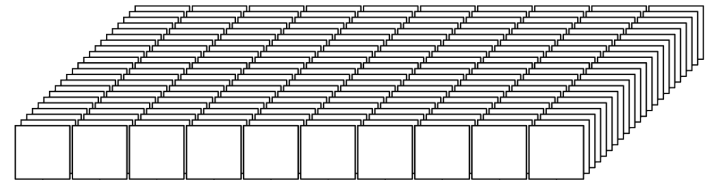
Séquences

N	E	V	F	H	T	G	I	K	Q
---	---	---	---	---	---	---	---	---	---

Objets réels

Embeddings

Encodages numériques



"Objet augmenté"
doté d'"expertise"

aucune interprétabilité

Allons-nous demander à un expert de nous expliquer son expertise?

Q: avons-nous besoin d'interpréter les résultats d'une architecture d'apprentissage profond ?

2. L'IA nous permet de verifier ce que devient important pendant l'apprentissage

Nous pouvons verifier **les poids** de cet apprentissage

Ces poids sont le fruit d'un calcul colossale, difficilement reproductible ! Les poids constituent une vrai valeur pour l'entreprise.

Q: les poids des modèles doivent-ils être rendus publics ou non ?

Commentaire 1 - L'IA en science est puissante — Comment communiquer ses résultats

Nous publions aujourd'hui des prédictions que le public **ne comprend pas** dans leur intégralité.

Avec AlphaMissense, on peut désormais consulter en ligne les niveaux de pathogénicité associés aux mutations dans nos gènes. Si nous connaissons notre génome, nous pouvons vérifier si certaines mutations peuvent altérer le fonctionnement de nos gènes. **Le risque d'une interprétation erronée de ces valeurs de pathogénicité demeure.**

Il est essentiel de réfléchir aux usages, pas uniquement aux performances. La communauté scientifique a un rôle fondamental à jouer dans cette régulation.

Q: Comment rendre public ses résultats ? Comment les accompagner ?

Commentaire 2 - L'IA nous remplacera un jour ?

Les **mécanismes de raisonnement** seront développés aussi.

Mais pour cela il faut que les machines apprennent à "utiliser" les données **pour en créer** d'autres **plus complexes**. Cela nous amènera vers l'implémentation de l'imagination!

Q: Si une machine peut imaginer, qu'est-ce qui reste propre à l'esprit humain ?

Q: Sommes-nous prêts à faire confiance à une machine qui invente ce qu'elle ne connaît pas encore ?