

**IA:
fake généralisé
ou
technologie cognitive appropriée**

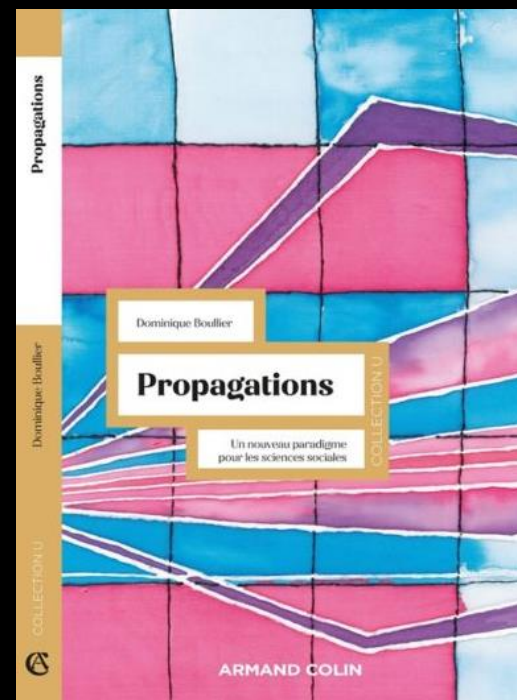
Dominique BOULLIER

SciencesPo

Biennale du numérique
Enssib Lyon
Novembre 2023

Survivre aux passages à l'acte pour penser les IA

- Le coup de force de Open AI, disrupteur en chef: un cas réussi de diffusion/propagations forcée
- = Penser l'IA **malgré** Cheat GPT et pouvoir choisir (merci!) et non s'adapter
- Sans oublier qu'il existe:
 - d'autres IA que celles d'Open AI
 - d'autres IA que génératives,
 - d'autres IA que connexionnistes,
 - d'autres IA que propriétaires,
 - d'autres IA que conversationnelles,
 - d'autres IA que généralistes
 - d'autres IA que gratuites
 - d'autres IA que boites noires
 - d'autres IA que la version actuelle (ChatGPT4- Turbo)



La double face des IA génératives: interfaces et architecture de connaissances

- Une technologie cognitive: tout dispositif matériel/logiciel qui nous fait penser différemment (agency)
 - ex: un panneau indicateur/ une carte/ un GPS (couplé à l'action, couplé à un environnement, et distribué
- Deux dimensions associées :
- Agent conversationnel (chat): l'accès aux données, les commandes et l'interaction (prompts et conversation)
- Large Language Models: modèle de langage pré-entraîné avec GPT3 ou 4 ou avec d'autres transformeurs sur des données en masse



Ce que font (et nous font faire) les interfaces

- Les interfaces sont seulement la surface?
 - Oui précisément et nous n'avons accès -entre humains comme avec les machines- qu'à la surface et aux signaux qui sont transmis
- Les interfaces sont le cadre cognitif de perception des signaux et d'envoi des commandes: tout le reste est opaque. L'agentivité est attribuée par les humains à la machine ou pas.
- Puissance du test de Turing: est intelligente une machine qui paraît intelligente aux yeux des humains.

Revisiter l'histoire des interfaces: mutations sensorielles

Interface	1980-1984	1990-1991	2001	2007	2023
Système	WYSIWYG GUI	Web Liens hypertexte	Wikipédia écriture wiki	Iphone tactile	<i>ChatGPT texte + vocal</i>
Modalité sensorielle	Vue + toucher (souris)	Vue + commande clic	Vue+ clic+ écrit	Vue+ commande tactile écran	<i>Commande vocale</i>

Autres modalités dans le domaine du jeu (non
passées dans les usages ordinaires):
Les manettes (dont la Wii), les pédales, les tapis
Les lunettes de Réalité Virtuelle

Revisiter l'histoire des interfaces: mutations industrielles

Interface	1980-1984	1990-1991	2002	2007	2023
Système	WYSIWYG GUI	Web liens hypertexte	Wikipédia écriture wiki	Iphone tactile	<i>ChatGPT texte +vocal</i>
Modalité sensorielle	Vue + toucher (souris)	Vue + commande clic	Vue+ clic+ écrit	Vue+ commande tactile écran	<i>Commande vocale</i>
Rupture avec système précédent	Les lignes de code	La bibliothèque	L'encyclopédie	La commande par touches et codes	<i>Commande par fonctions</i>
Effet industriel	Le PC (le Mac)	Le Web	Le Web 2.0	Le smartphone Les applis et les appstores	<i>Les agents conversa- tionnels de l'IA générative</i>

Revisiter l'histoire des interfaces

Mutations des affordances cognitives

Interface	1980-1984	1990-1991	2002	2007	2023
Système	WYSIWYG GUI	Web Liens hypertexte	Wikipédia écriture wiki	Iphone tactile	<i>ChatGPT texte + vocal</i>
Organisation des connaissances	OS et accès aux logiciels	Graphe de liens hypertextes à explorer avec moteur de recherche	Écriture distribuée, révisable, coopérative, temps réel	Apps fermées (la fin du web)	<i>Réponse à tout non sourcée générée par IA probabiliste</i>
Processus attentionnel	Fabriquer des documents comme un imprimeur	Surfer (explorer, rechercher)	Contribuer	Être alerté, réagir	<i>S'immerger dans l'interaction sans friction</i>

Non pris en compte: le Peer to peer et la distribution des fichiers

Revisiter l'histoire des interfaces: encastées dans un design organisationnel

Interface	1980-1984	1990-1991	2002	2007	2023
Système	WYSIWYG GUI	Web Liens hypertexte	Wikipédia écriture wiki	Iphone tactile	<i>ChatGPT texte + vocal</i>
Design organi- sationnel	Passage à l'informatique personnelle sans réseau	Connectivité des documents sourcés par des URL	Distribution des contributions, P2P	OS mobiles, appstores avec rente d'accès API	<i>Délégation à UNE firme d'IA</i>
Dérives	Vers les stations de travail pro (GPU)	Vers le moteur de recherche qui devient moteur de réponses	Vers les plateformes publicitaires des réseaux sociaux	Vers l'alerte permanente et la traçabilité généralisée	<i>Vers l'outil de réponse totale immédiat</i>

La double face des IA génératives: interfaces et architecture de connaissances

- L'interface conversationnelle nécessite et donne accès à UN modèle de langue (généraliste ou spécialisé)
- Ses propriétés encapsulent un modèle de connaissance implicite qui est perçu comme savant voire intelligent
- L'interface ne donne aucun accès
 - au « modèle conceptuel » (Norman) qui permettrait de comprendre le fonctionnement des LLM et d'ajuster ses attentes
 - ni aux conditions de son développement, dont l'entraînement des données par des humains esclaves (Pretrained)

Un modèle vectoriel et probabiliste qui exige toujours plus de ressources

- Large Language Models = Une vectorisation de la langue, avec statistiques et probabilités pour générer des séquences prédictibles: zéro sémantique!
- Tokenisation généralisée (d'où les milliards de paramètres) avec réduction à la lettre ou à la phrase (effet des Transformeurs: augmentation de l'attention, pour maintenir la mémoire de l'environnement)
- Set de données toujours plus vaste pour l'apprentissage pour les versions généralistes: très consommateur d'énergie, une approche extractiviste antisobriété
- Mais attention beaucoup d'IA génératives spécialisées et nourries par les sets de données des experts. Avec des modèles plus réduits et moins exigeants en ressources, et même des IA génératives coopératives (hestia.ch)

Une architecture de connaissances encapsulée dans les GPT

- Réponses et non recherche (à tout prix même celui des hallucinations)
- Absence de sources
- Immédiateté
- Correction stylistique et adaptation extrême aux environnements et à leur prérequis: une IA moyenne et conformiste par définition .
- Illusion du tout : le monde est entre nos mains en temps réel

Grok is designed to answer questions with a bit of wit and has a rebellious streak, so please don't use it if you hate humor!

A unique and fundamental advantage of Grok is that it has real-time knowledge of the world via the X platform. It will also answer spicy questions that are rejected by most other AI systems.

Une portée subjective et émotionnelle non contrôlée

- Un jeu (Baricco The Game): qui joue avec qui?
- Une personnalisation extrême qui dissout les distances cognitives et affectives: le cas de « Project December. Simulate the Dead » qui était seulement textuel
 - article The Conversation D. Boullier + R. Alfonso Romero
- **Counterfeit people** : les fausses personnes (Dennett: aussi criminel que la fausse monnaie)
 - la perte de la confiance
- Tout l'opposé d'un système de connaissances à prétention scientifique: auteurs (et non générateurs), sources, discussion, référencement et indexation, exploration, validation.

Les évolutions en cours: toujours plus spéculatif et perversif

- Interface vocale (grand public et plus)
- Couplage avec le moteur de recherche= plus de fiabilité? Ou plus d'intoxication?



- Orchestrateur : une connaissance-action (demande de services avec accès aux API): réservation de billet
- ChatGPT4-Turbo: générateur de GPT personnels adaptés avec place de marché (coordonnées par OpenAI) pour que chacun monétise les services ainsi conçus dans son domaine

Un régime de vérité en péril

- Œuf + défunt + hallucinations + immédiateté = extension du domaine du fake (extended deepfake)
- Une nouvelle maxime conversationnelle: fais en sorte de ne jamais croire que les maximes conversationnelles et conventionnelles de Grice sont respectées par l'IA: quantité, qualité (vérité), pertinence, clarté
- Violation (ou floutage) permanente des maximes comme hypothèse raisonnable.... Mais invivable! (nécessaire en régime totalitaire)

Des interfaces toujours plus intuitives pour des modèles toujours plus opaques



- Séduction des promesses à travers les interfaces
- Incapacité critique et obligation de s'adapter
- Acceptation de l'opacité

Gagnant / gagnant ???

- Le cauchemar des scientifiques



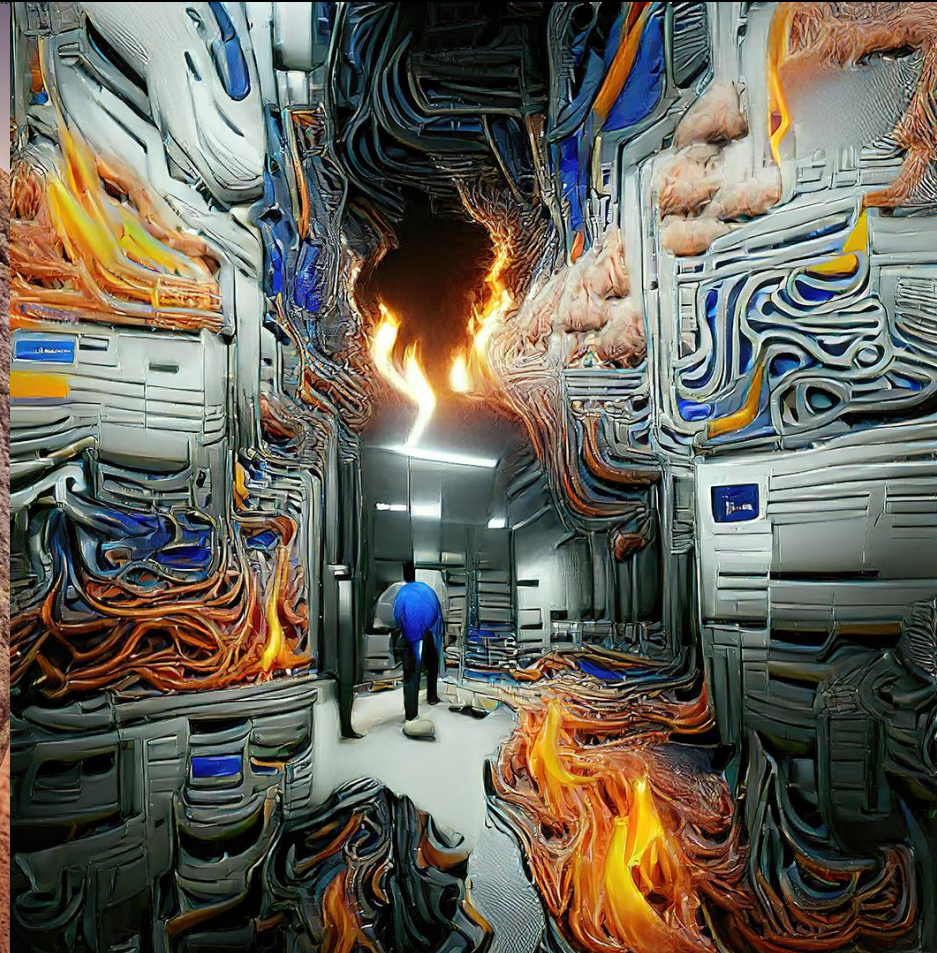
Gagnant / gagnant ???

- Le désarroi des publics



Gagnant / gagnant ???

- Le bonheur des artistes (??)



Gagnant / gagnant ???

- La tentation des politiques



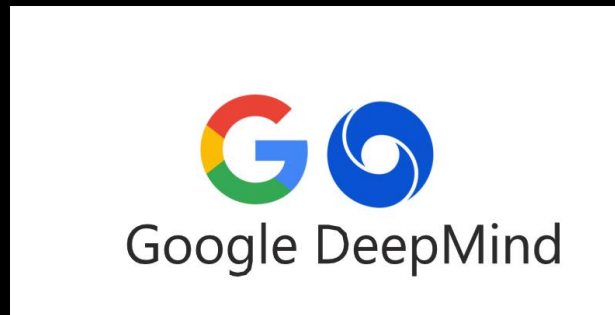
Gagnant / gagnant ???

- La jouissance des disrupteurs (et de Poutine!)



Gagnant / gagnant ???

- Les profits des plateformes



Gagnant / gagnant ???

- Le cauchemar des scientifiques
- Le désarroi des publics
- Le bonheur des artistes (??)
- La tentation des politiques
- La jouissance des disrupteurs (et de Poutine!)
- Les profits des plateformes

- Quelle est l'arène de décision?
-
- Quelles sont les alternatives? Open source? Datasets publics?
IA explicable? Autorisation de mise sur le marché?

