

Perspectives et défis

Résumé

Le document explore les perspectives et défis de l'intelligence artificielle (IA), en mettant l'accent sur la transition vers l'**IA Générale (AGI)**, une IA capable d'égaliser les capacités cognitives humaines. Il aborde les différentes approches du développement de l'IA, entre **modélisation symbolique** et **apprentissage connexionniste**, et discute des avancées technologiques récentes, notamment dans le traitement du langage naturel et la vision par ordinateur.

L'AGI est présentée comme une opportunité majeure, notamment pour l'Europe qui bénéficie d'une recherche de pointe dans ce domaine. Toutefois, elle pose des questions fondamentales sur la conscience, la prise de décision et l'impact potentiel sur la société, notamment sur l'emploi et les interactions humaines.

Le document met également en lumière les **cinq grands obstacles** au développement de l'AGI :

- Le mur de la confiance** (problèmes de transparence et d'explicabilité des décisions de l'IA).
- Le mur de la sécurité** (risques liés à la manipulation des données et aux deepfakes).
- Le mur de l'énergie** (consommation énergétique excessive des modèles d'IA avancés).
- Le mur de l'interaction avec l'humain** (difficulté de comprendre et d'interpréter les intentions humaines).
- Le mur de l'inhumanité** (manque de raisonnement causal et de sens commun dans les IA actuelles).

Enfin, le document examine les **dangers potentiels** d'une IA autonome, notamment son influence sur les décisions humaines, son impact sur le marché du travail et le développement de robots humanoïdes capables de surpasser les humains intellectuellement et physiquement. Il souligne la nécessité d'une régulation proactive et d'une approche éthique pour garantir que l'IA reste un outil bénéfique plutôt qu'une menace.

Terme	Définition
IA Générale (AGI)	Intelligence artificielle capable d'imiter toutes les capacités cognitives humaines, incluant la prise de décision et l'apprentissage autonome.
Apprentissage connexionniste	Approche basée sur des réseaux de neurones artificiels imitant le fonctionnement du cerveau humain.
Modélisation symbolique	Approche de l'IA basée sur des règles logiques explicites pour la prise de décision.
Deepfake	Contenu audio, vidéo ou image manipulé par IA pour imiter des personnes réelles de manière trompeuse.
Traitement du langage naturel (NLP)	Domaine de l'IA permettant aux machines de comprendre et de générer du texte en langage humain.
Vision par ordinateur	Technologie permettant aux machines d'analyser et d'interpréter des images ou vidéos.
Explicabilité de l'IA	Capacité d'un modèle d'IA à justifier ses décisions de manière compréhensible pour les humains.
Open Source	Modèle de développement où le code d'un logiciel est librement accessible et modifiable.
Mur de l'inhumanité	Limite actuelle des IA qui ne possèdent pas de raisonnement causal, de sens commun ou de conscience.
Robot humanoïde	Machine conçue pour imiter la morphologie et les interactions humaines.