Intelligence artificielle, deepfakes et crise de confiance

Peut-on encore croire ce que l’on voit, lit et entend ?

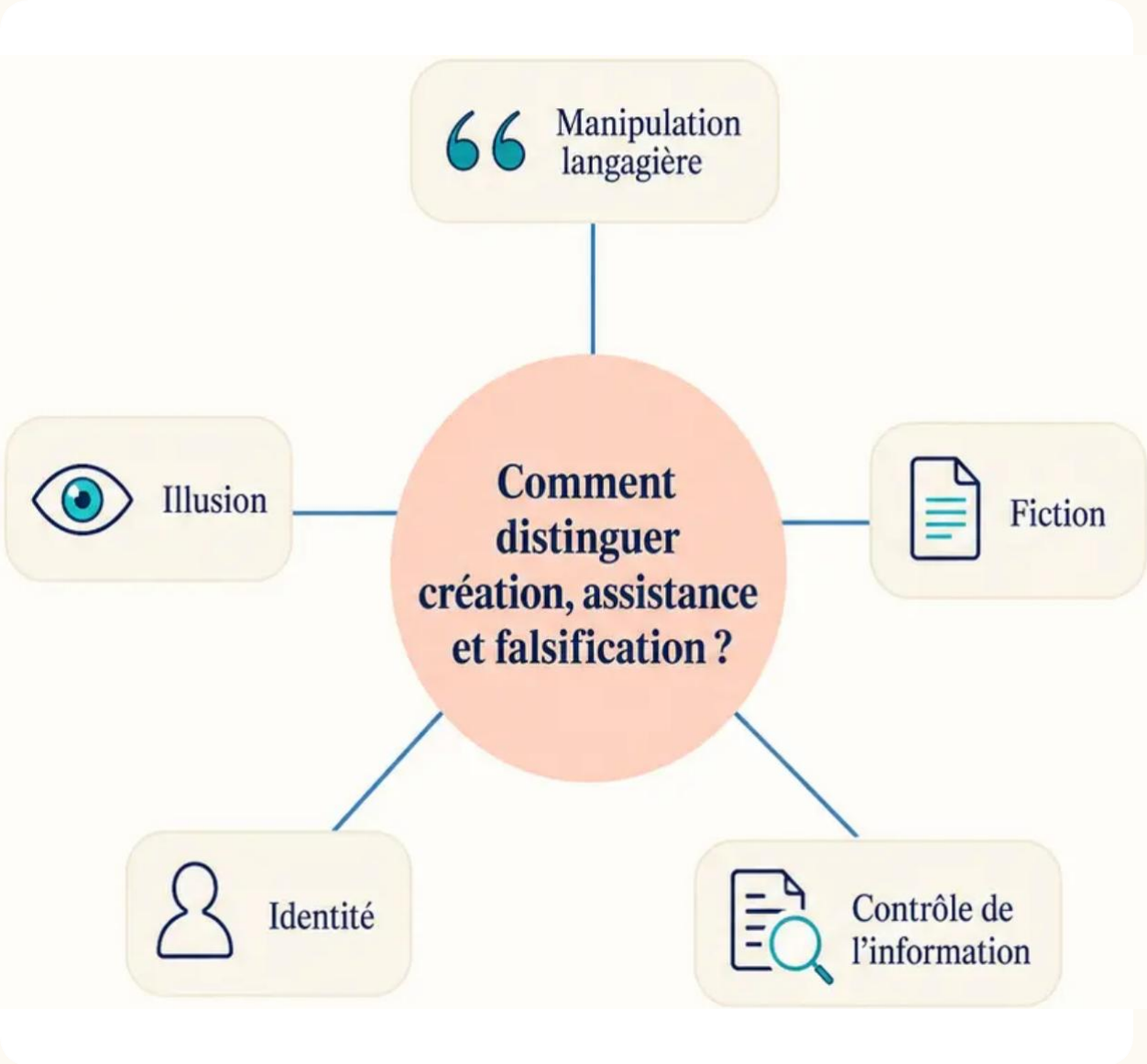
Culture générale et expression — BTS 2e année



Une technologie qui concentre les tensions du thème

Objectifs du chapitre

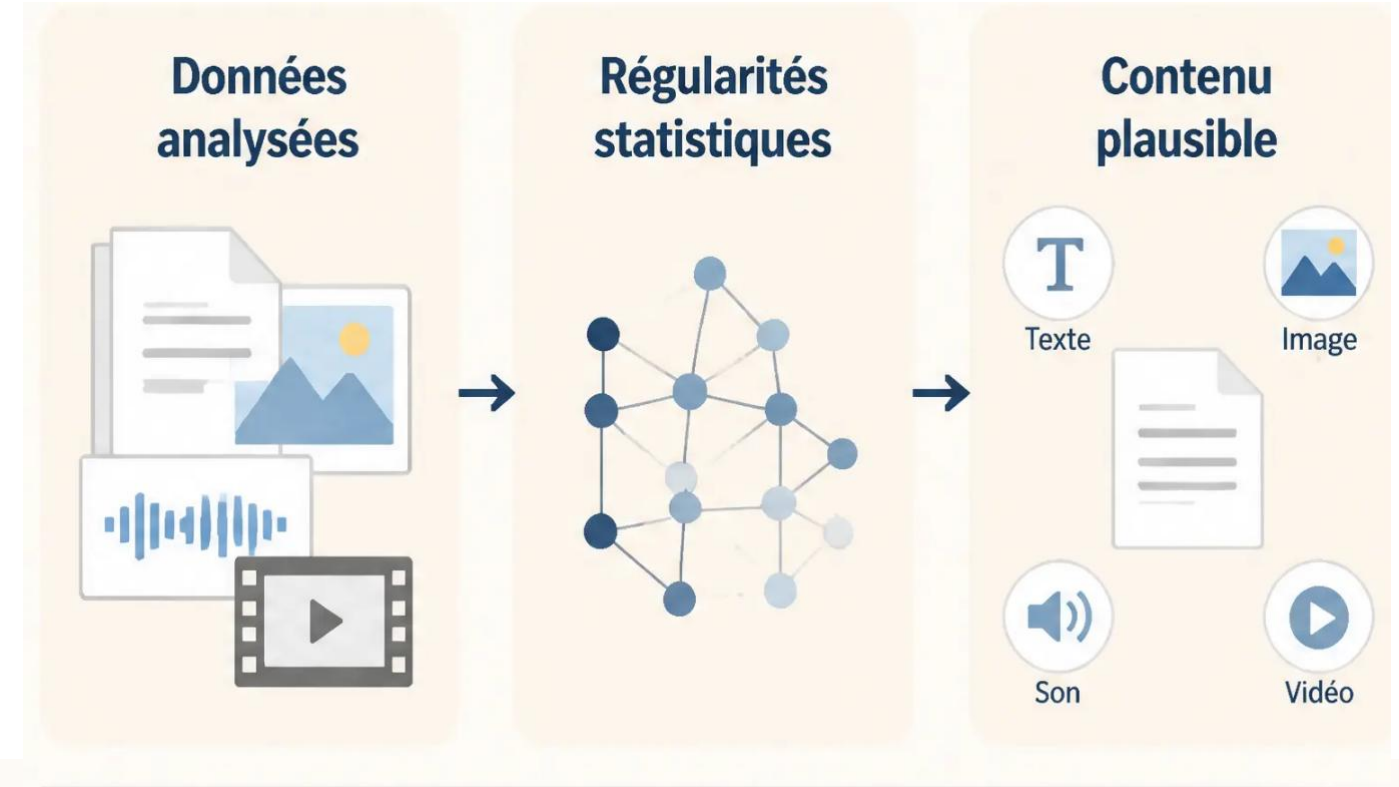
- Comprendre ce qu'est un contenu généré
- Étudier les deepfakes et l'authenticité
- Réfléchir à la confiance à l'ère de l'IA



L'IA prolonge des questions déjà rencontrées dans l'année.

Qu’est-ce que l’intelligence artificielle générative ?

Des systèmes capables de produire des contenus originaux — textes, images, sons et vidéos — à partir de l’analyse statistique de grandes quantités de données.



Un contenu plausible n’est pas automatiquement un contenu vrai.

L’IA reconnaît des régularités statistiques.

Elle ne comprend pas nécessairement le monde et ne garantit pas la vérité.

Création, assistance ou falsification ?

Analyser le statut d'un contenu généré ou modifié.

	💡 Création	🔧 Assistance	📄 Falsification
Quelle est l'intention ?			
Le caractère généré est-il signalé ?			
Cherche-t-on à faire passer le contenu pour un document authentique ?			





Activité : textes, images ou voix synthétiques — à fournir par l'enseignant.

Objectif : distinguer l'usage, l'intention et la prétention à l'authenticité.

Les hallucinations : cohérent ne signifie pas vrai

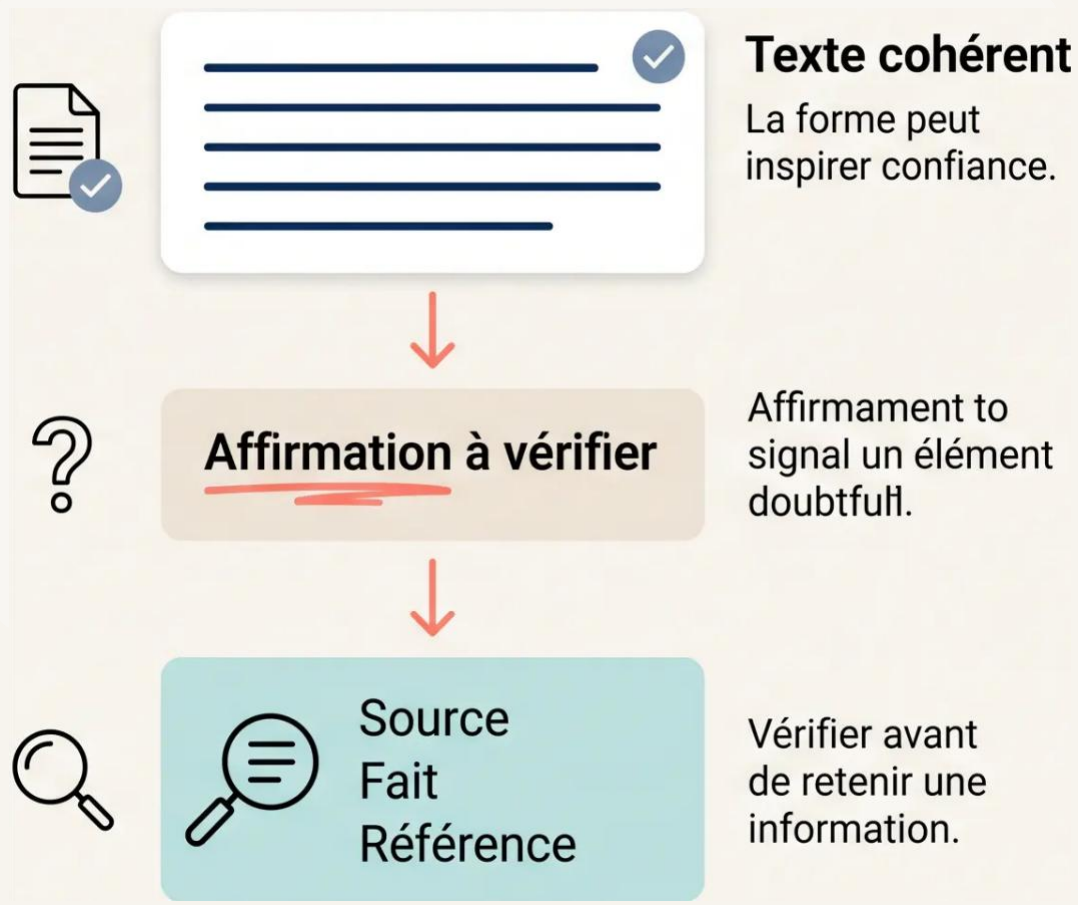
Une erreur structurelle

Une hallucination est une affirmation fausse produite avec une apparente confiance. Elle ne résulte pas d'une intention de tromper, mais des limites du système.

- Le texte peut être fluide et bien structuré.
- Il peut mêler informations exactes et inventions.
- La continuité du style rend l'erreur difficile à distinguer.

La qualité de la forme ne suffit pas à valider le contenu.

Apparence crédible ≠ information validée



Deepfakes : quand l’imitation du réel devient difficile à détecter

Des contenus audiovisuels où une personne réelle est remplacée ou imitée par une reconstruction artificielle, rendant la manipulation difficile à détecter.

Politique

Fausses déclarations attribuées à des personnalités réelles.

Vie privée

Contenus non consentis.

Information

Difficulté à authentifier les images et vidéos servant de preuves.



La détection reste possible avec des compétences adaptées et des outils qui progressent.

Étude de cas à fournir par l’enseignant

Image, son, texte : que peut-on encore tenir pour une preuve ?

Comparer les supports, puis vérifier à partir des documents fournis.

Texte

Peut être cohérent, bien structuré et pourtant faux.



Image

L'image générée fragilise la preuve photographique, déjà affectée par la retouche.



Son / vidéo

Les contenus synthétiques interrogent l'authenticité audiovisuelle.

Tableau de vérification — à compléter en classe

	Ce que je peux encore croire	Ce que je dois vérifier
Texte		
Image		
Son / vidéo		

Supports à fournir par l'enseignant

La crise de confiance et le dividende du menteur

Les deepfakes fragilisent la confiance dans les preuves… y compris lorsqu'elles sont authentiques.

Un faux peut paraître vrai



Manipulation

Un contenu fabriqué peut sembler crédible.

Crise de confiance

Dividende du menteur

Un document vrai peut être rejeté



"Fabriqué par IA"

Une personne prise en faute peut nier une preuve réelle.

Le "dividende du menteur" : la possibilité de discréditer une preuve authentique en affirmant qu'elle a été fabriquée par IA.

Le doute généralisé peut profiter davantage aux menteurs qu'aux victimes de manipulation.

Technique, usages et responsabilité

La technique est-elle responsable?

Distinguer les causes, les choix et les relais avant de prendre position.

1

Capacité technique

Quelde l'outil permett à être produit ?

Contenu généré

2

Intention de l’utilisateur

Examine le choix de créer, modifier ou décait?

Conditions de diffusion

Des canaux, bleus, contexte, et l’échelle de circulation

4

Responsabilité du relais

Que fait-on avant de partager ou de croire un contenu?

5

Arguments : la technique a une part de responsabilité

- —
- —
- —

Arguments : la responsabilité dépend des usages

- —
- —
- —

Des réponses pour reconstruire la confiance

Des solutions techniques, institutionnelles et éducatives,
dont le déploiement généralisé prendra du temps.



Détection

Des outils progressent pour repérer les contenus manipulés.



Watermarking

Marquage invisible des contenus générés.



Certification C2PA

Standard de certification de l'authenticité, développé notamment par Adobe.



Cadre et éducation

Règlement européen sur l'IA adopté en 2024 ; renforcement de l'éducation aux médias.

Solutions émergentes d'authentification

Point de vigilance : aucune solution isolée ne suffit ; les réponses techniques demandent du temps pour être déployées largement.

Atelier — élaborer une charte d’usage raisonné

En groupes, complétez une charte pour un usage responsable des contenus générés par IA.

Notre charte d’usage raisonné

Signaler le recours à un contenu généré	
Distinguer création, assistance et falsification	
Vérifier les affirmations et leurs sources	
Examiner l’authenticité des textes, images, sons et vidéos	
Respecter le consentement	
Assumer sa responsabilité lors de la diffusion	



Travail en groupes



Produire des règles concrètes et compréhensibles



Supports à fournir par l'enseignant

SIGNALER VÉRIFIER RESPECTER ASSUMER

Synthèse et entraînement à l’essai

Sujet d’essai

L’intelligence artificielle menace-t-elle la vérité ?



Construire une réponse argumentée et nuancée.

Préparer sa réponse

1 Analyser les termes du sujet

2 Confronter risques et contre-mesures

3 Mobiliser les notions du chapitre

4 Construire une argumentation nuancée

Notions à mobiliser

contenus plausibles, hallucinations, deepfakes, preuve photographique

crise de confiance, dividende du menteur, détection, certification, éducation aux médias

À retenir

L’IA amplifie des tensions anciennes entre réel et représentation.

Elle fragilise l’authentification, sans rendre toute vérification impossible.

L’esprit critique, la vérification des sources, la culture générale et des institutions fiables permettent de reconstruire la confiance.