

Algorithmes et data: quels enjeux pour le service public audiovisuel ?

Antonin Descampe
antonin.descampe@uclouvain.be



Audition dans le cadre du renouvellement du contrat de gestion RTBF
Parlement de la Fédération Wallonie-Bruxelles
25 mai 2022



Quelles sont les technologies les plus susceptibles de jouer un rôle important dans les 5 prochaines années dans les médias ?

Data from a survey of 234 media leaders in 43 countries for Journalism, Media and Technology Trends and Predictions 2021.

AI

69%

5G

18%

New devices

9%

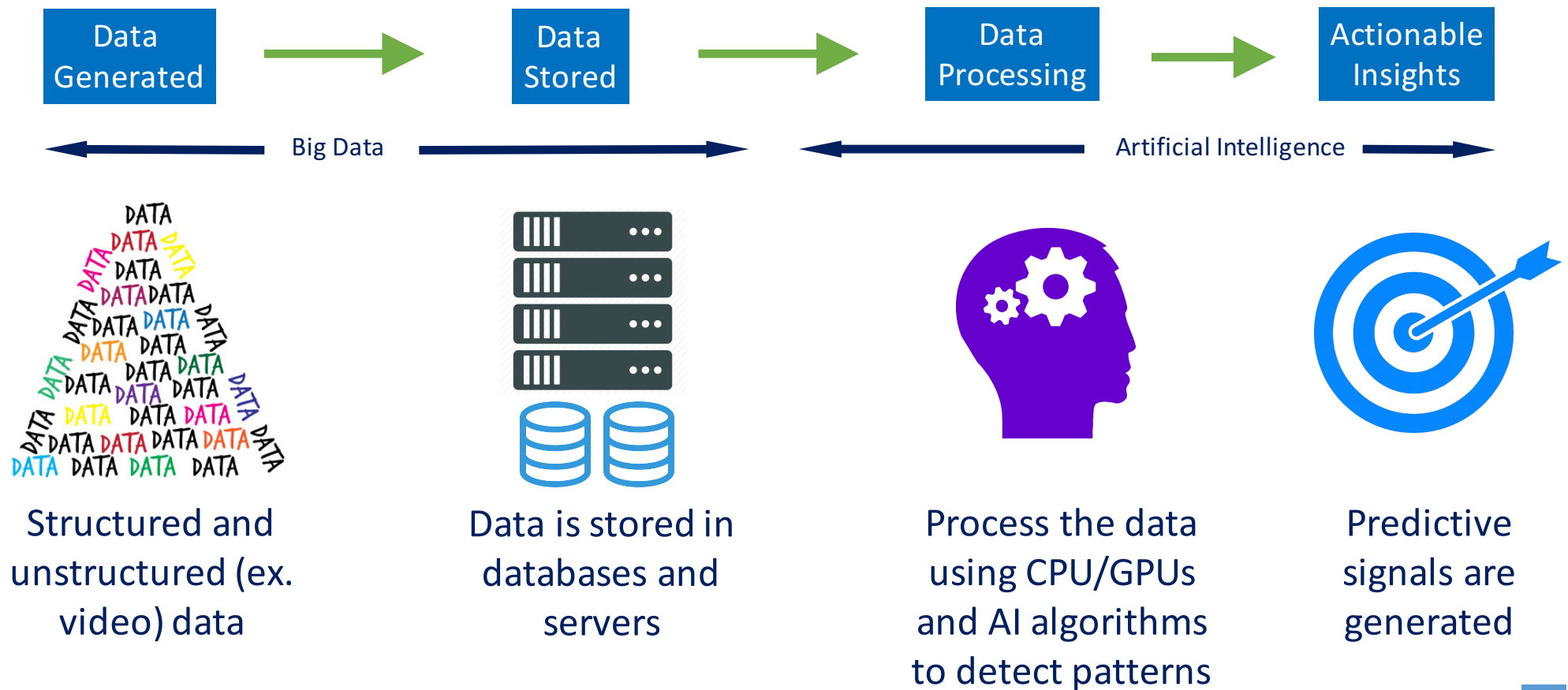
Don't know

4%

Q19. In your view, which of the following enabling technologies is likely to be most important for media companies to understand over the next five years? n=227



IA = Décisions automatiques prise à partir de l'analyse de grandes quantités de données



Médias & automatisation

veille
/ collecte

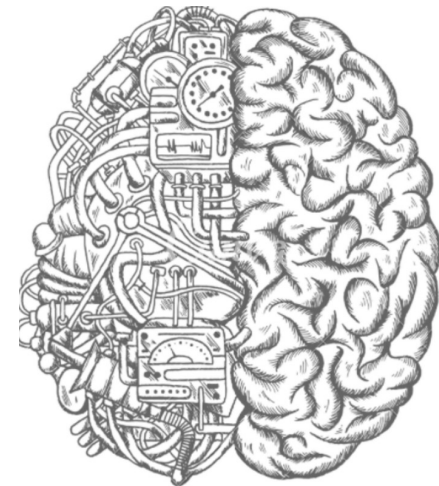
- Story discovery
- **Data mining**
- Social media monitoring
- Fact checking

Production

- **Text generation**
- Multi-medias formats from text
- Efficient instantaneous translations
- Data visualization, summarization

Distribution
/ réception

- Comment / content moderation
- **Personalized recommendation**
- Conversational newsbots
- Headline testing



Le **traitement automatisé** des données touche toutes les étapes de **collecte**, **production** et **réception** de l'information, et affecte tant les aspects **éditoriaux** que ceux liés au **marketing** ou au **business model** d'un media

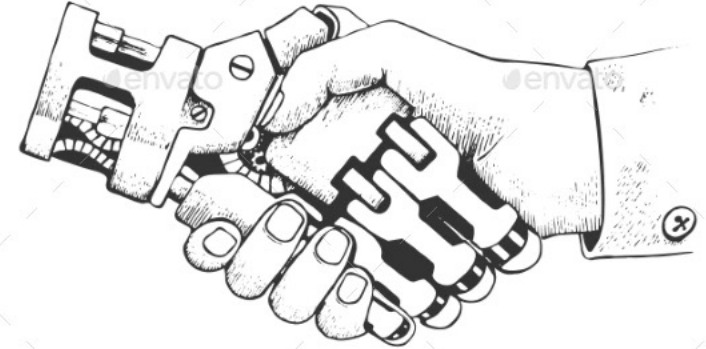


Constat 1: Evolution de l'écosystème des médias

Les métiers de l'information
et des médias vont-ils disparaître ?

Non !

- Humains et machines sont doués pour des choses différentes
 - **Machines**: filtrer, associer, classer, prioriser de (très) **grandes quantités de données**, (très) **rapidement** (*scale, speed, personalization*)
 - **Humains**: intelligence sociale (écouter, négocier, convaincre, expliquer, contextualiser), compréhension des causes, pensée experte, ...



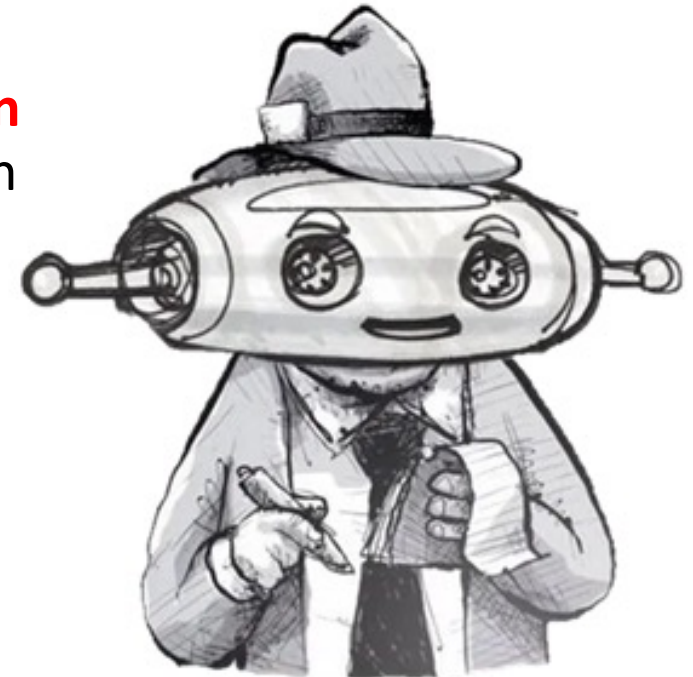
Vers une **hybridation**, une **complémentarité** croissante
entre travail humain et travail automatisé.



Constat 1: Evolution de l'écosystème des médias

Nécessité de

- ... un **investissement massif dans l'innovation** et l'automatisation de la chaîne de production de l'information
- ... consacrer une grande partie de cet investissement à
 - la **formation** (compétences numériques, prototypage de formats d'information, *computational thinking*) et la **diversification des profils**
 - une **co-construction de cette hybridation** avec les professionnels des médias, les journalistes, les start-ups, pour une innovation pérenne, acceptable pour les professionnels et en phase avec les valeurs et les pratiques du métier (approche **écosystémique** et **interdisciplinaire**)





Constat 2: Omniprésence des décisions automatiques

ALGORITHMS ALREADY INFLUENCE:



WHO WE TALK TO

ALGORITHMS IN SOCIAL MEDIA SELECT WHOSE POSTS WE SEE, SHAPING OUR RELATIONSHIPS



WHAT WE BUY

SUGGESTION ENGINES ('PEOPLE ALSO BOUGHT') SHAPE AWARENESS AND INFLUENCE PURCHASES



WHERE WE GO

THE ROUTES SELECTED BY DIGITAL MAPS DETERMINE THE NEIGHBOURHOODS WE VISIT AND BUY PROPERTY IN



HOW MUCH WE EARN

STOCK TRADING SYSTEMS DETERMINE SHARE PRICES, IMPACTING OUR SAVINGS AND OUR SALARIES



WHO WE MARRY

ALGORITHMS IN DATING APPS PLAY A KEY ROLE IN DETERMINING WHO WE MEET, DATE, AND MARRY



Automatisation = enjeu démocratique majeur

- Quel **modèle de société** promouvoir à travers ces décisions automatiques ?
 - Objectif poursuivi par les algorithmes de recommandation : maximiser les interactions ?
 - Bulles de filtres
 - Polarisation accrue de la société
- Quels **mécanismes de contrôle** de ces décisions ? Même conçus de manière vertueuse, les algorithmes sont biaisés
 - Biais des concepteurs
 - Biais des données (en particulier les données socio-numériques)
 - Causalité ≠ corrélation



Une **automatisation « de service public »** est possible (bonne nouvelle !)

- **Personnalisation** permise par les algorithmes n'est pas mauvaise en soi : absurde de s'en passer car peut être mise au service des valeurs de service public
 - Accès large à une information fiable qui rende les gens plus autonomes
 - Pluralisme
 - Cohésion sociale (contre la polarisation du débat public)
 - Promotion des libertés fondamentales
 - Fonctionnement de la démocratie
- Une **collecte massive de données** est nécessaire pour un fonctionnement optimal
 - ... et peut se faire de manière transparente et vertueuse, en laissant l'utilisateur en contrôle

Devoir d'exemplarité et d'excellence du service public



Des algorithmes vertueux, explicables et auditables

Des algorithmes, munis des garde-fous nécessaires, peuvent œuvrer en faveur d'un engagement des **publics moins affinitaires** et lutter **contre la polarisation du débat public** et le **désordre informationnel**

- **Transparence et explicabilité** : informer et expliquer en termes simples
- **Auditabilité** : audits réguliers et publics pour déceler les biais éventuels
- **Métriques appropriées** : temps de lecture plutôt que nombre de clics
- **Sérendipité et découvrabilité des contenus** : favoriser des contenus hors de la sphère de préférence de l'utilisateur
- **Vérification de l'information** : déployer et développer Faky par l'automatisation
- **Segmentation plus fine des publics** : assouplir et affiner la segmentation actuelle des publics

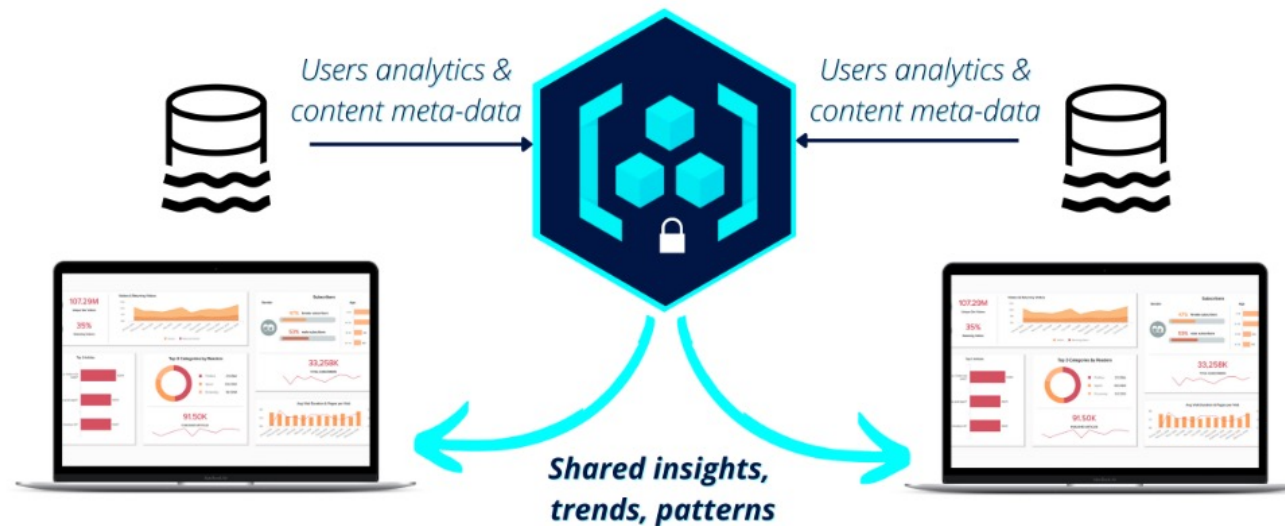




Une gestion des données transparente, collaborative et (pt-à-pt) décentralisée

En matière de données, ne pas se tromper d'ennemi : les concurrents sont les grandes plateformes Meta, Spotify, Netflix, etc.

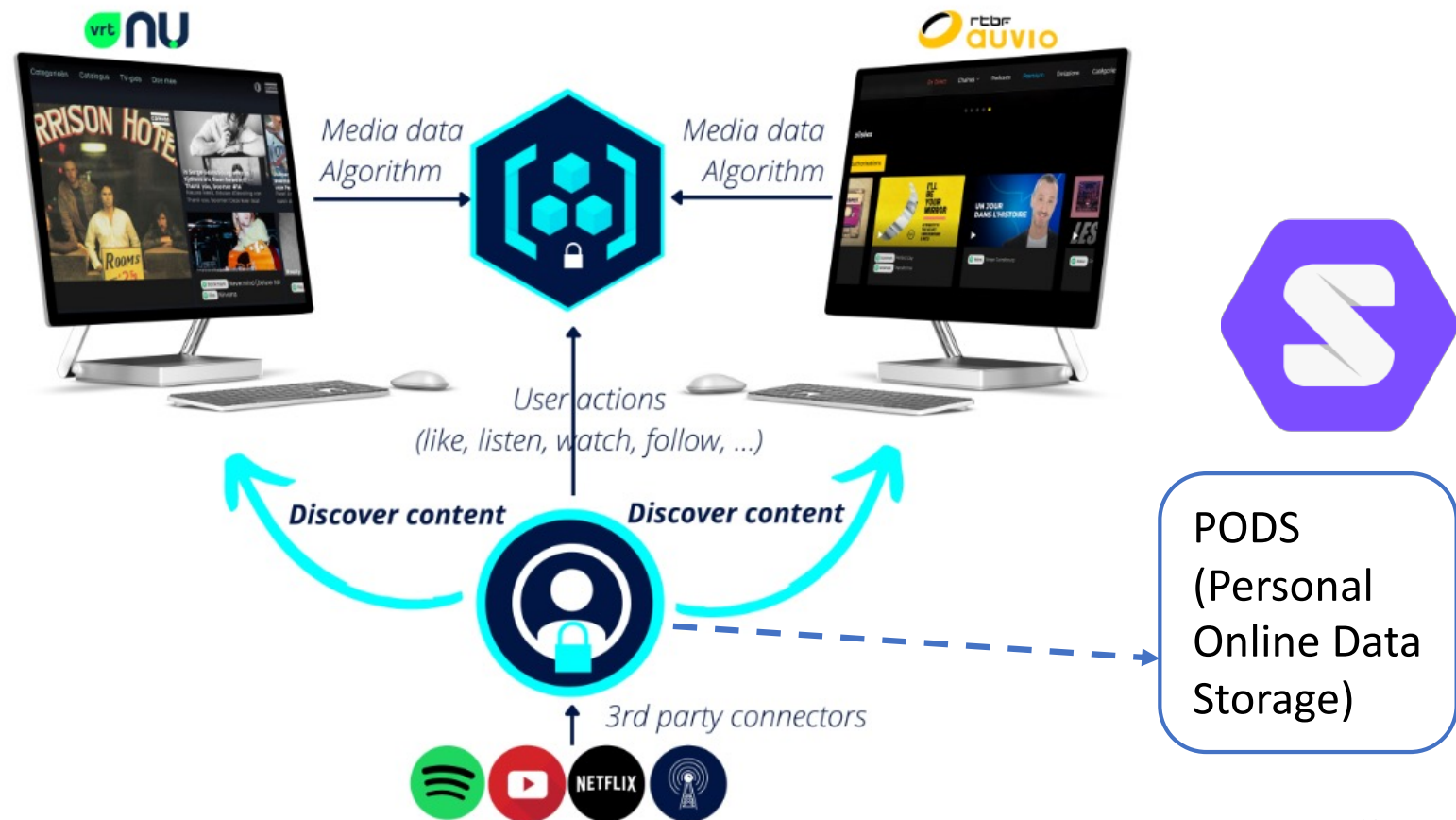
- Encourager la collaboration par la **mise en commun de *data assets*** (et d'*algorithmic assets*) via un service d'intermédiation
- Cette collaboration renforce la **résilience des acteurs locaux** face aux plateformes, tout en garantissant la propriété des *assets*.





Une gestion des données transparente, collaborative et (pt-à-pt) décentralisée

L'enjeu est, pour la RTBF, mais plus largement pour la FWB, de ne pas rater le tournant du **web3** et de la **décentralisation** des données





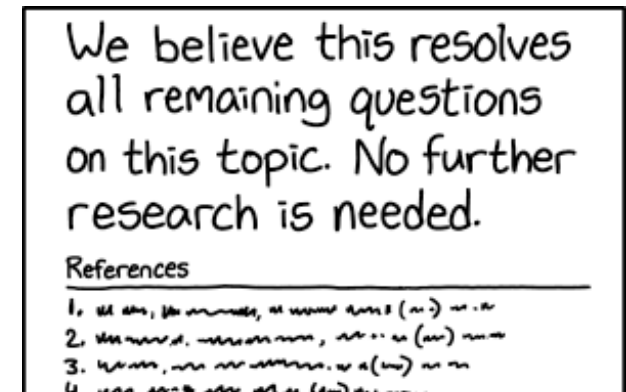
Une gestion des données ouverte à des fins de recherche scientifique

Les données de la RTBF (articles, vidéos, images, sons, analytics) sont un matériau unique et précieux pour des recherches académiques sur l'évolution de la société et du rapport aux médias

- Les questions de genre et de diversité dans les articles
- Les sujets traités et la priorité qui leur est donnée
- L'évolution de la mise en récit de différents problèmes publics au cours du temps.
- L'évolution du poids des différentes catégories d'information (politique, société, culture, sport, santé...) au fil du temps dans les différents médias
- *etc*

Contribuer utilement à l'avancement de la science dans plusieurs disciplines est en phase avec les valeurs de service public et bénéficierait également à la RTBF

→ Mise à disposition des données à des fins de recherche



JUST ONCE, I WANT TO SEE A RESEARCH PAPER WITH THE GUTS TO END THIS WAY.



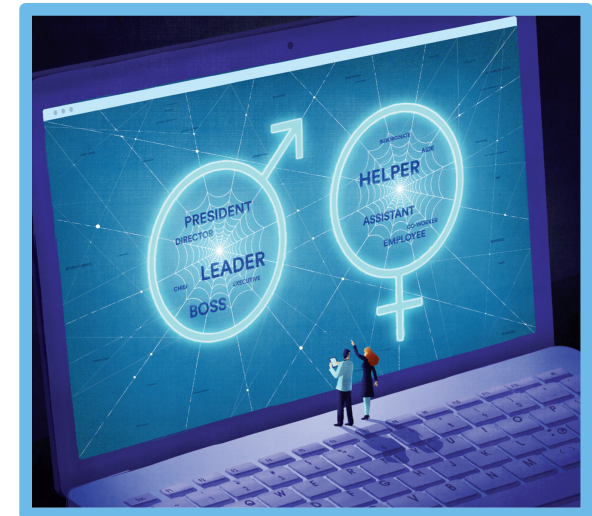
Exemple de projet de recherche « Social AI »

Objectifs :

Utiliser l'intelligence artificielle et la linguistique pour analyser les biais présents dans les articles de presse francophone.

A l'aide de la technique des "word embeddings", qui transforme les mots d'un corpus en vecteurs sur base de leurs contextes d'apparition.

Observer l'évolution de la représentation des femmes dans la presse belge francophone avant/après la révolution *#metoo*.





Exemple de projet de recherche « Social AI »

Algèbre vectorielle

Bruxelles – Belgique + France = ... **Paris**

PS – gauche + droite = ... **MR**

princesse – femme + homme = ... **prince**

prince – homme + femme = ... **épouse**

atelier – homme + femme = ... **cuisine**

Différents types d'**opérations mathématiques** peuvent être réalisées sur les vecteurs.

Opérations réalisées sur les vecteurs générés à partir de 60.000 articles RTBFINFO (2013-2015)



Exemple de projet de recherche « Social AI »

Evolution des biais

On peut mesurer l'évolution des rapports entre différents mots grâce à la **distance vectorielle**.

Si distance = 1, mots totalement similaires

Si distance = -1, mots totalement contraires

Distance entre le mot <i>femme</i> et ...	Articles 2013-2015		Articles 2020-2021
<i>foyer</i>	0.21	↘	0.07
<i>ménage</i>	0.27	↘	0.17
<i>pouvoir</i>	-0.02	↗	0.03
<i>ministre</i>	-0.04	↗	0.05
<i>PDG</i>	-0.08	=	-0.07
<i>journaliste</i>	0.22	=	0.22



En conclusion

Take-home messages

1. Investir dans la formation et un écosystème complet d'innovation
2. Développer des algorithmes vertueux, transparents, explicables, auditables et favorisant la découvrabilité de contenus
3. Evoluer vers un modèle décentralisé de gestion de données et s'unir avec d'autres acteurs locaux afin de faire le poids face aux grandes plateformes socio-numériques
4. Ouvrir les données à des fins de recherche scientifique sur l'évolution de la société, et l'objectivation de la couverture médiatique en Belgique francophone.

Merci de votre attention



Références

- Bogaert, J., Descampe, A., de Marneffe, M.-C., and Standaert, F.-X. (2022, June). Automatic and manual detection of generated news: Case study, limitations and challenges. *To Appear in the Proceedings of the ACM International Workshop on Multimedia AI against Disinformation (MAD 2022)*.
- Carlebach, M., Cheruvu, R., Walker, B., Magalhaes, C. I., and Jaume, S. (2020). News Aggregation with Diverse Viewpoint Identification Using Neural Embeddings and Semantic Understanding Models. *Proceedings of the 7th Workshop on Argument Mining*, 59–66.
- Coddington, M. (2015). Clarifying journalism’s quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3(3), 331–348.
- Descampe, A., Massart, C., Poelman, S., Standaert, F.-X., and Standaert, O. (2021). Automated news recommendation in front of adversarial examples and the technical limits of transparency in algorithmic accountability. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-021-01159-3>
- Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media* (Harvard University Press).
- Garg, N., Schiebinger, L., Jurafsky, D., and Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16), E3635–E3644. <https://doi.org/10.1073/pnas.1720347115>
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ArXiv:1301.3781 [Cs]*. <http://arxiv.org/abs/1301.3781>
- Rouvroy, A. (2017). La vie n’est pas donnée. *Etudes Digitales*, 2(2), 195–217.