Questions de la transparence et des rémunérations
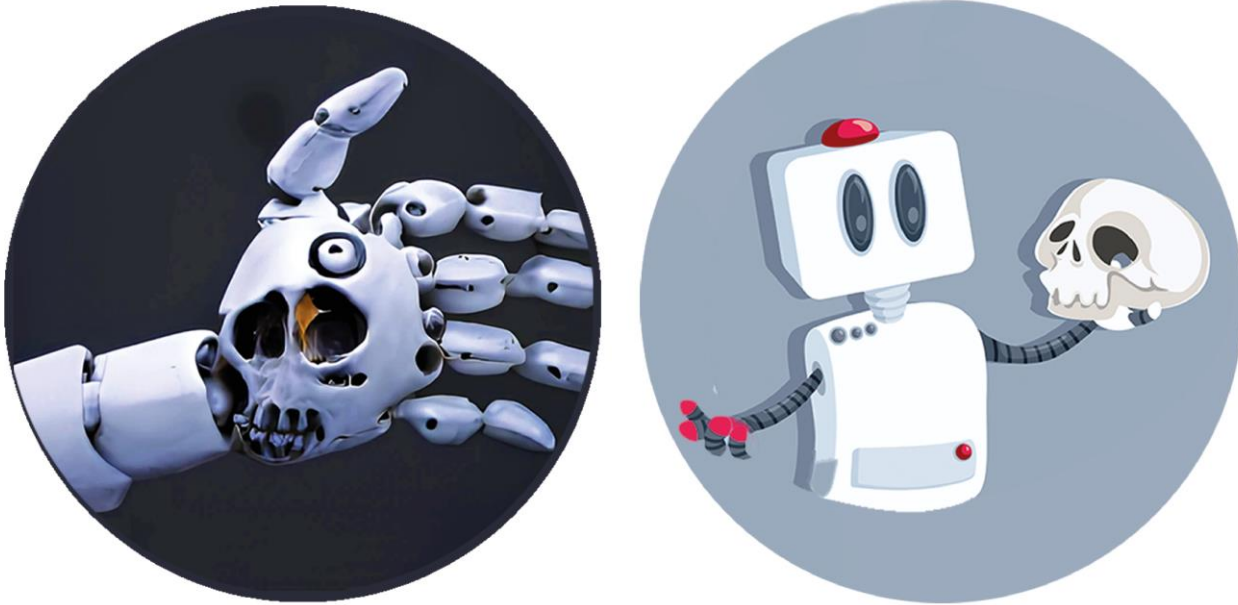

TO BOT OR NOT TO BOT

Nina George

- membre du l'ALAI Allemagne
- Écrivaine
- Commissaire aux affairs politique du European Writers' Council (EWC)

EWC European Writers' Council

Questions de la transparence et des rémunérations
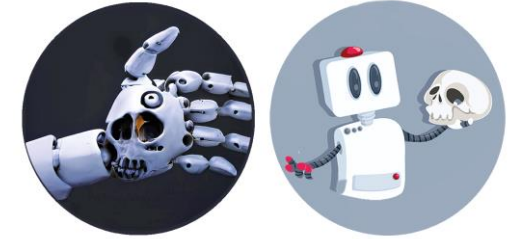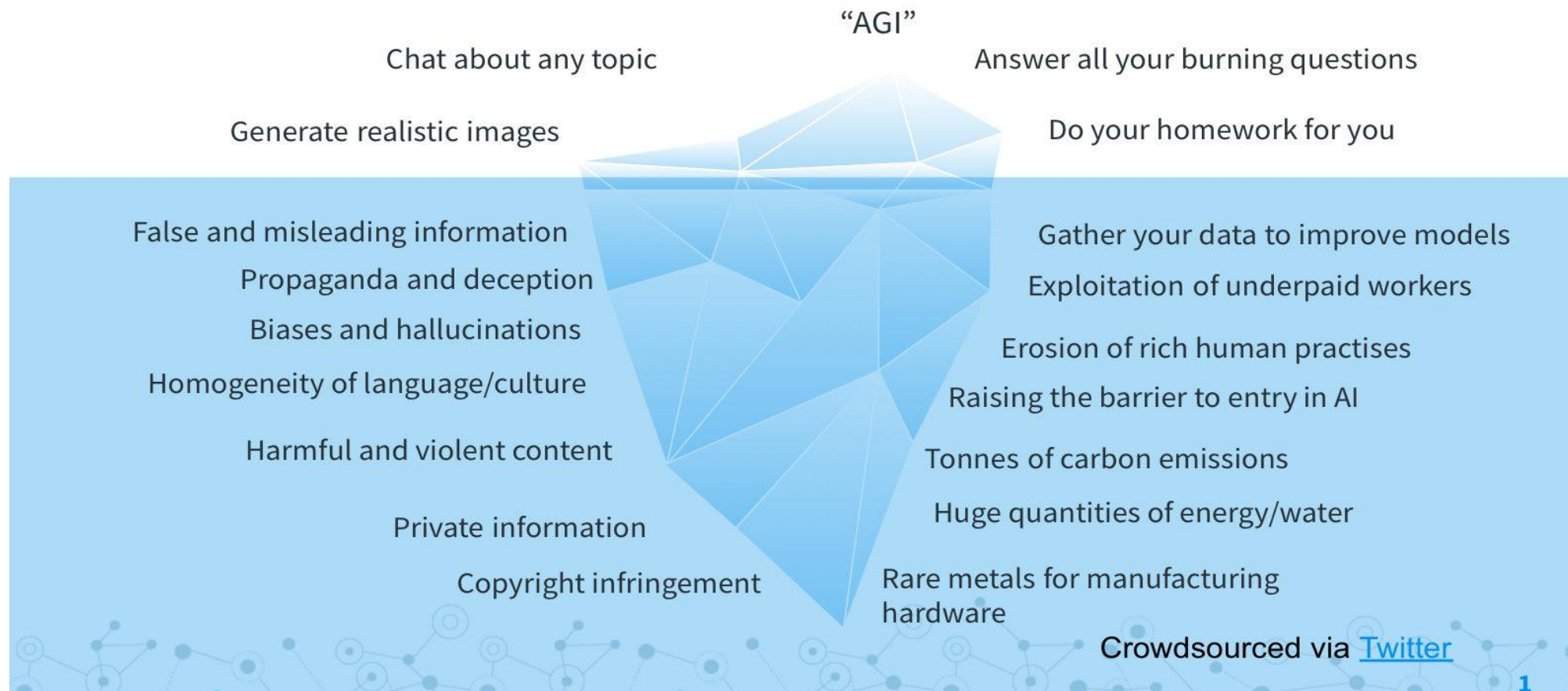
table of content

- Pre-Amble
- LLM in the Book Sector
- Transparency and Input
- Transparency and Output
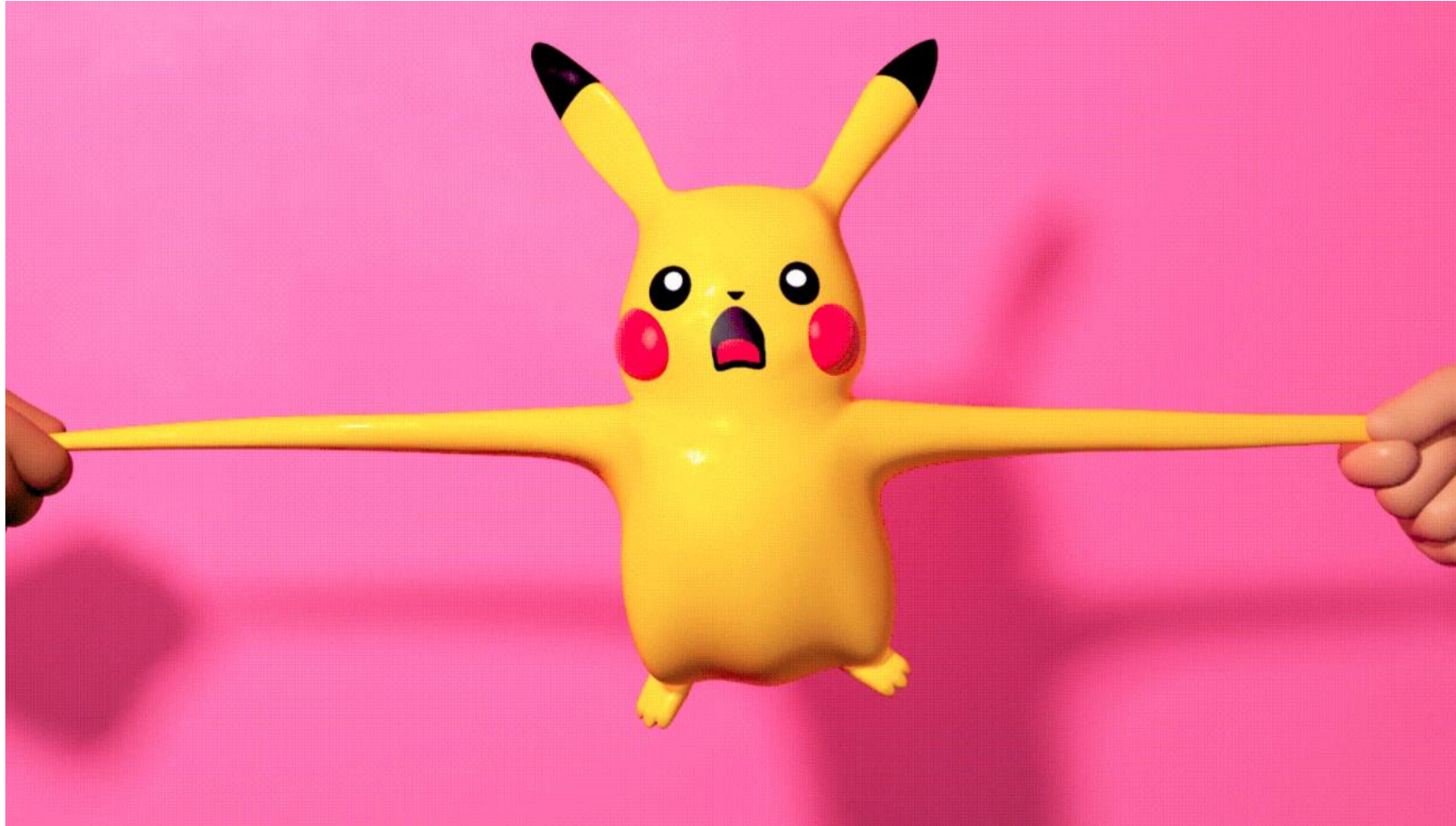- Solutions, anyone?

Sommaire de présentation

Pré-Amble

LLM dans le secteur du livre

Transparence et input

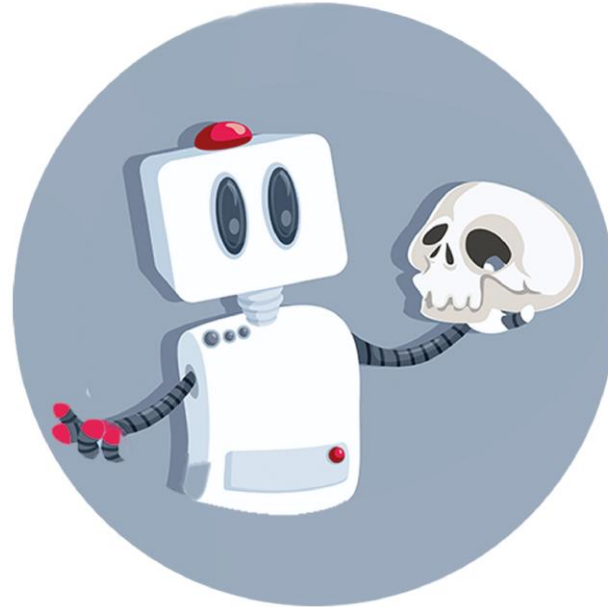Transparence et output

Besoin de solutions ?

TO BOT OR NOT TO BOT

Generative AI Models: Costs vs Benefits

"AGI"

Chat about any topic

Answer all your burning questions

Generate realistic images

Do your homework for you

False and misleading information

Gather your data to improve models

Propaganda and deception

Exploitation of underpaid workers

Biases and hallucinations

Erosion of rich human practises

Homogeneity of language/culture

Raising the barrier to entry in AI

Harmful and violent content

Tonnes of carbon emissions

Huge quantities of energy/water

Private information

Rare metals for manufacturing hardware

Copyright infringement

Crowdsourced via Twitter

1

It's the Tamagotchi-Effect

AI in the Book Sector

TO BOT OR NOT TO BOT

…that is (no longer) the question.

Transparency on the Input: Piracy rulez

```
The Little Bride - Anna Solomon
The Little Buddhist Monk & the - Cesar Aira
The Little Butterfly - Sherry Shahan
The Little Colonels Knight Comes Riding_nodrm
The Little Communist Who Never - Lola Lafon
The Little Edges - Fred Moten
The Little French Bakery Cookbook - Susan Holding
The Little French Bistro - Nina George
The Little French Cookbook - Murdoch Books Test Kitchen
The Little Friend - Donna Tartt
The Little Girl Who Was Too Fond of Matches
The Little Green Spoon - Indy Power
The Little Grey Men - BB
The Little Guide to Butterflies - Tom Frost
The Little Guide to Leaves - Tom Frost
The Little Guide to Vintage Sho - Melody Fortier
The Little Hotel - Christina Stead
The Little Immigrants - Kenneth Bagnell
The Little Kitten - Judy Dunn
The Little Library - Kim Fielding
The Little Library Cookbook - Kate Young
The Little Library Year - Kate Young
The Little Magazines - Ian Hamilton
The Little Man - Liza Alexandrova-Zorina
The Little Minister - J. M. Barrie
The Little Money Book
The Little Moroccan Cookbook_nodrm
The Little Old Lady Behaving Ba - Catharina Ingelman-Sundberg
The Little Old Lady Who Broke A - Catharina Ingelman-Sundberg
The Little Old Lady Who Struck Lucky Again - Catharina Ingelman-Sundberg
The Little Paris Bookshop - Nina George
The Little Pasta Cookbook - Murdoch Books Test Kitchen
The Little Pizza Cookbook - Murdoch Books Test Kitchen
```

Dataset of over 184.000 titles. Sources: Bit Torrent Piracy Sites, e.g. Library Genesis, Z-Library, further sources.

Among them:

- 31 books by **Salman Rushdie**
- **98 by Stephen King**
- **48 by John Grisham**
- **63 different editions of Harry Potter Books**
- **48 books by the Dalai Lama**
- 2 by **Nina George**
- 2 by Sebastian Fitzek (translations and originals)

- The Books were published before the TDM exception was valid, and not licenced
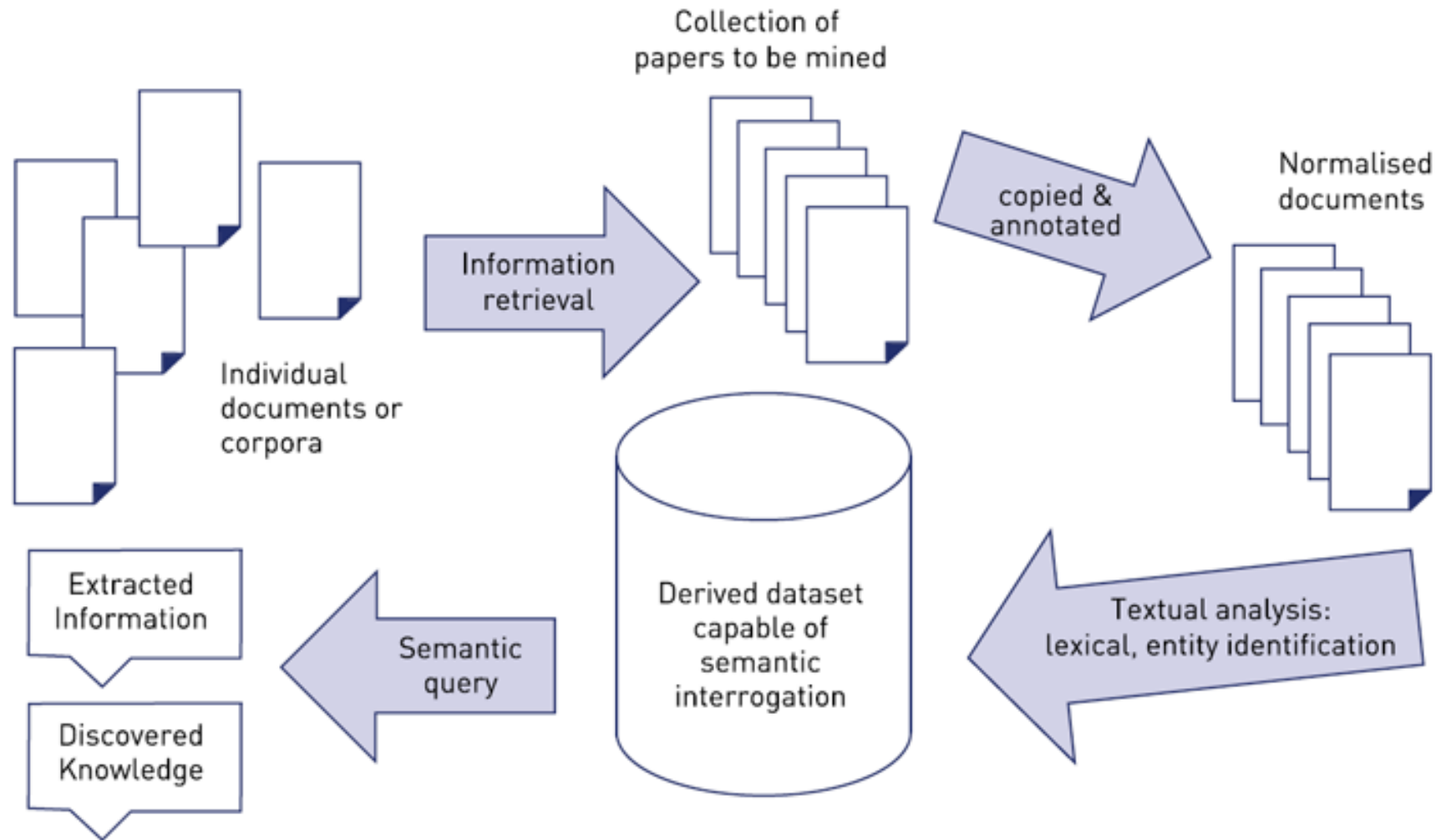- Double Copyright infringement

Transparency on the Input: The Stanford check

Grading Foundation Model Providers' Compliance with the Draft EU AI Act

Source: Stanford Research on Foundation Models (CRFM), Institute for Human-Centered Artificial Intelligence (HAI)

Draft AI Act Requirements	GPT-4	Cohere Command	Stable Diffusion v2	Claude	PaLM 2	BLOOM	LLaMA	Jurassic-2	Luminous	GPT-NeoX	Totals
Data sources	●○○○	●●●○	●●●●	○○○○	●●○○	●●●●	●●●●	○○○○	○○○○	●●●●	22
Data governance	●●○○	●●●○	●●○○	●●○○	●●●○	●●●●	●●○○	○○○○	○○○○	●●●○	19
Copyrighted data	○○○○	○○○○	○○○○	○○○○	○○○○	●●●○	○○○○	○○○○	○○○○	●●●●	7
Compute	○○○○	○○○○	○○○○	○○○○	○○○○	●●●○	●●●○	○○○○	●○○○	●●●○	17
Energy	○○○○	●○○○	●●○○	○○○○	●●●○	●●●○	●●●○	○○○○	○○○○	○○○○	16
Capabilities & limitations	●●●●	●●●○	●●●○	●●●○	●●●●	●●●○	●●○○	●●●○	●●○○	●●●○	27
Risks & mitigations	●●●○	●●●○	●●○○	●●○○	●●●○	●●●○	○○○○	●●○○	●○○○	○○○○	16
Evaluations	●●●○	●●●○	●○○○	○○○○	○○○○	○○○○	●●○○	●●●○	○○○○	●●○○	15
Testing	●●●○	●●●○	○○○○	○○○○	○○○○	○○○○	○○○○	●●○○	○○○○	○○○○	10
Machine-generated content	●●●○	●●●○	●●●○	●●●○	●●●○	●●●○	●●●○	●●●○	●●●○	●●○○	21
Member states	●●○○	●●○○	●●○○	○○○○	●●○○	●●○○	○○○○	○○○○	○○○○	○○○○	9
Downstream documentation	●●●○	●●●○	●●●○	○○○○	●●●○	●●●○	●●●○	○○○○	○○○○	●●●○	24
Totals	25 / 48	23 / 48	22 / 48	7 / 48	27 / 48	36 / 48	21 / 48	8 / 48	5 / 48	29 / 48	

TDM exception: does is cover machine learning?

Collection of papers to be mined

Information retrieval

copied & annotated

Normalised documents

Individual documents or corpora

Textual analysis: lexical, entity identification

Derived dataset capable of semantic interrogation

Semantic query

Extracted Information

Discovered Knowledge

Source: Cambridge Analytics

Solutions?

© Sally's Baking Addiction

Questions de la transparence
et des rémunérations

TO BOT OR NOT TO BOT

Merci!

EWC European Writers' Council