

Signs for Ethical AI:

A Route Towards Transparency

Dario Garcia Gasulla
dario.garcia@bsc.es

Atia Cortés
Sergio Alvarez-Napagao
Ulises Cortés

A historical parallelism

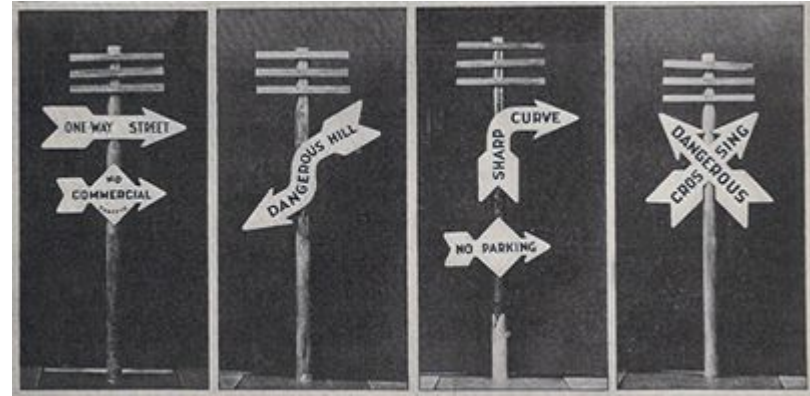
1900 vs 2020

A tale from 1900's

- ❖ New disruptive technology: **Automobiles**

- ❖ At the beginning

- A few benefit a lot
- Most benefit marginally



- ❖ Interaction between humans and technology needed regulation

- Registration, licenses, limits...
- Met with resistance. *Took decades!*

Fast Forward to 2020

- ❖ New disruptive technology: **AI**
- ❖ At the beginning
 - A few benefit a lot
 - Most benefit marginally
- ❖ Interaction between humans and technology needed regulation
 - Registration, licenses, limits...
 - Met with resistance. *Let's hope it does not take decades*



Signs: First level of information

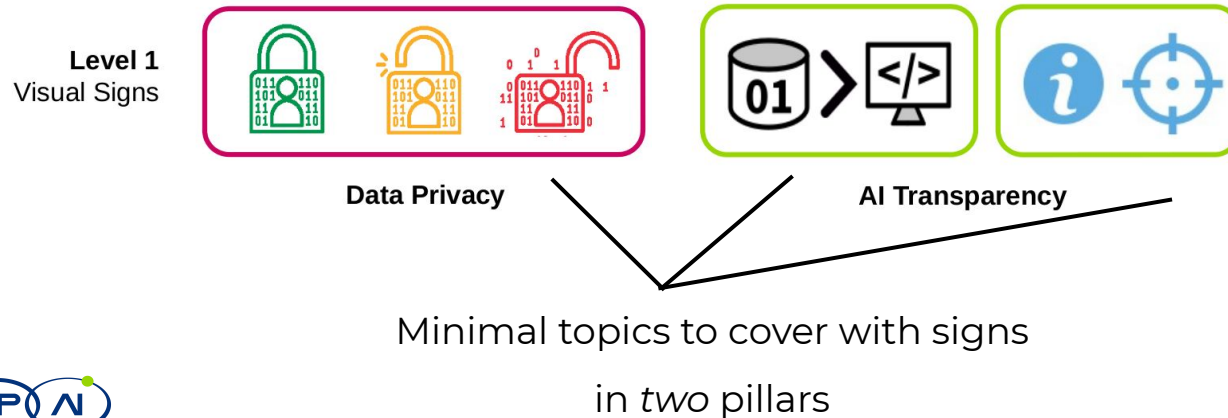
- ❖ Signs can be **great**
 - Complex ideas made simple (& non-intrusive)
 - Coarse boundaries
 - Minimal empowerment for all
- ❖ Signs can be **awful**
 - Detailed decision-making
 - Special cases/exceptions



Transparency Scheme

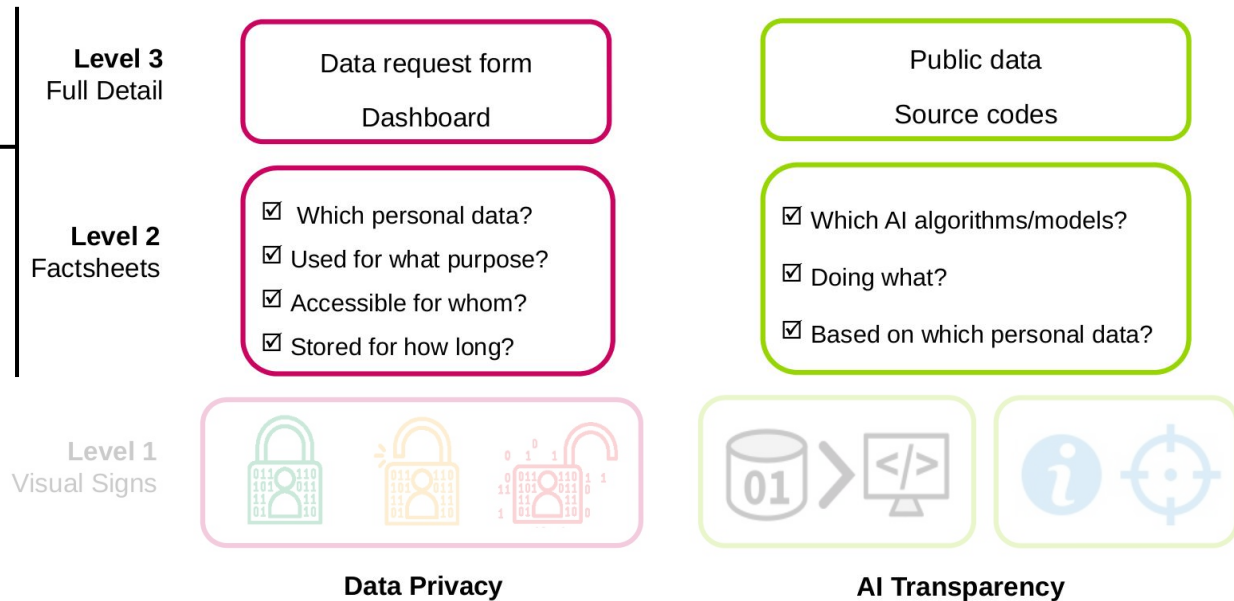
Three levels and Two pillars

Minimum Information



Covering all demands

Components
complementary
to signs
in *three levels*



Two pillars of information

❖ Data Privacy

- Key component of many AI
- Regulated (GDPR, PECR, ...)
- Current schemes (TCF) are annoying and confusing



❖ AI Transparency

- The basics of AI
- Unregulated
- We do not even know when there is an AI!



Three levels of information

❖ Level 1

- Access: All
- Goal: Minimum decision-making
- Medium: Visual signs



❖ Level 2

- Access: Interested
- Goal: Enable vast majority of policies
- Medium: Factsheets



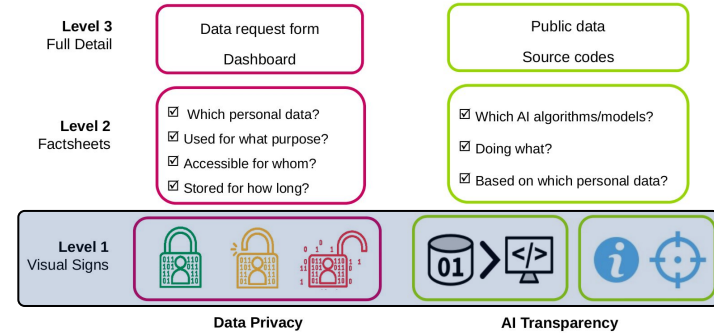
❖ Level 3

- Access: Experts
- Goal: Full decision process
- Medium: Raw data

A large, dense table of numbers, representing raw data. The numbers are arranged in a grid, with some cells containing negative values and others containing positive values. The table is oriented diagonally, with the top-left corner at the top of the image and the bottom-right corner at the bottom.

Level 1: Signs

What everyone ought to know



Signs of Data Privacy



No private data is gathered



Private data is gathered and used within the system



Private data is gathered and distributed to 3rd parties

❖ Basic decisions enabled

“Do I want this service to gather my data?”

“Do I want this service to distribute my data?”

Signs of AI Transparency



Data / Code of AI models is public



Code is public. Data is not.
Reproducibility? Bias?



Code and Data not public.
Opaque service.



Information is
independent of user



Information is **personalized**
and subjective to the user

- Manipulation
- Filter bubbles / echo chambers
- Partisanship
- Propaganda

◆ Basic decisions enabled

*“Do I want to use a service that
I cannot know what is doing?”*

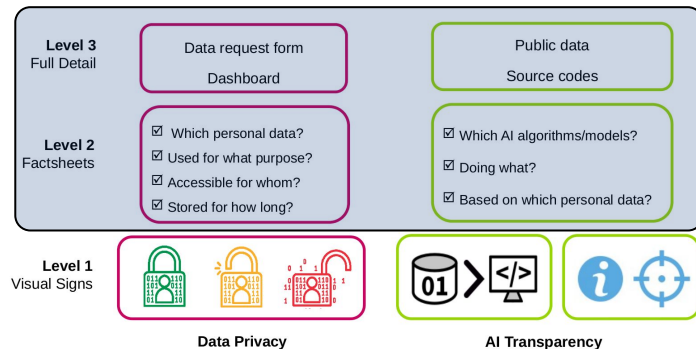
*“Do I want this service to be
aware of my identity?”*

Level 2: Factsheets

Level 3: Full Detail



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



Enabling *transparency by design*













Factsheets of Data Privacy (lvl2)

- ❖ GDPR guided
 - Which personal data?
 - Used for what purpose?
 - Accessible for whom?
 - Stored for how long?
- ❖ *Purpose & Data Type* refurbished
- ❖ Enables filters

Data Consent	Data Type	Purpose	Duration	Access
X	Personal Location	Chatbot	Less than 24 hours	Internal
	Query History	Chatbot	Less than 24 hours	Internal
✓	Navigation activity	Suggested News	Less than 1 month	Internal
X	Navigation activity	Personalized Ads	Less than 1 month	External Company A
✓	Use Statistics	Market Research	Less than 1 year	External company B

Factsheets of AI Transparency (lvl2)

- Which AI algorithms/models?
- Doing what?
- Based on which personal data?
- Is it gathering it?
- Can I know what it is doing?
- Is it subjective?

AI Consent	AI Purpose	Data Types	AI Service Conditions
 	Chatbot	Personal location, query history	   
 	Suggested news	Navigation activity	   

- ❖ Level 1 value is most restrictive value of Level 2
- ❖ Linked to other components

Full Detail (lvl3)

❖ Data Privacy

- Data Request Forms
- Consents and data dashboard

❖ AI Transparency

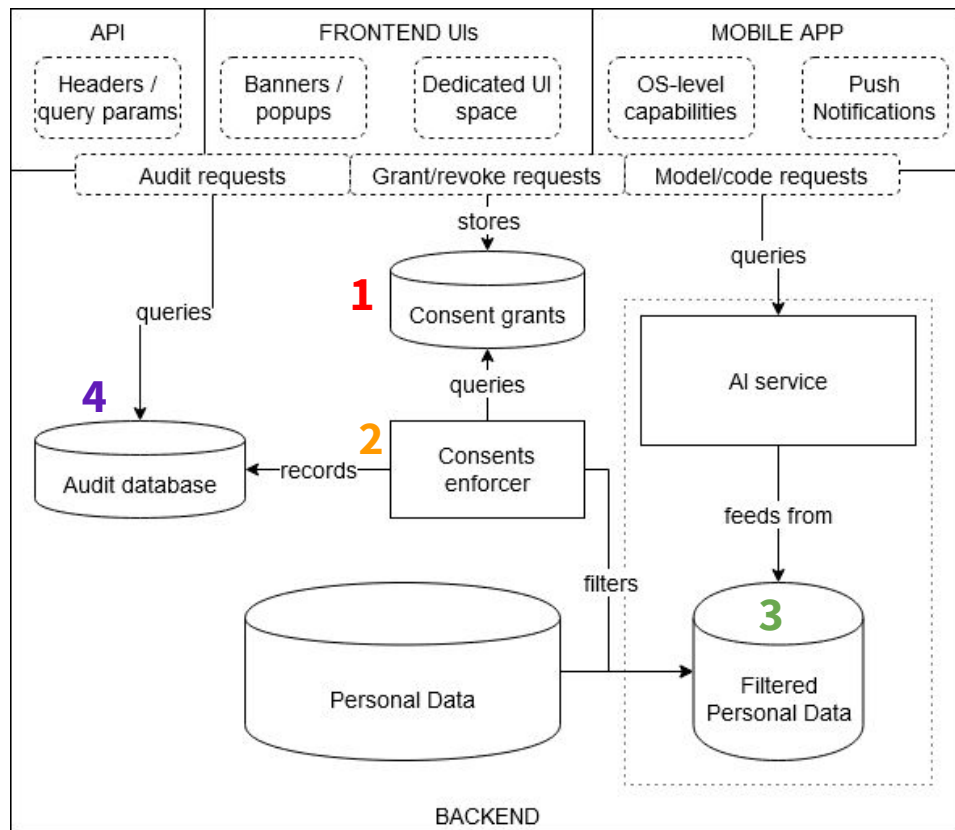
- Public data repositories
- Source codes and documentation

... the devil is in the details

- interface
- language
- delays
- accessibility

Transparency by Design

- ❖ Consents storage (snapshots) **1**
- ❖ Filter data based on consent **2**
- ❖ AI access only to filtered data **3**
- ❖ Record consents and activity **4**



Keep this in mind

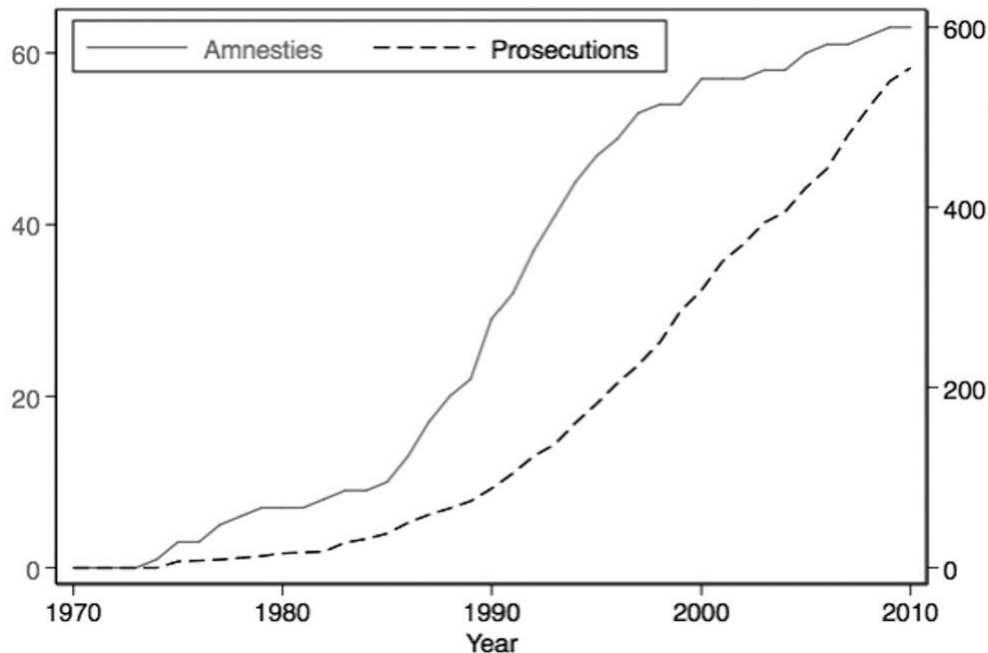
We can do better

Remarks

- ❖ Data Privacy rights are regulated but not implemented
- ❖ No balance of power between provider and consumer
- ❖ Transparency is power to the people
- ❖ User power is motivation for doing better
- ❖ Regulation will not work until the ethical principles are embraced

The case of Human Rights

- ❖ Regulation is not enough
- ❖ Universal Declaration of Human Rights by the UN General Assembly in 1948
- ❖ Ethics are not assumed and practiced overnight...
... but hopefully it does not take 50 years again



Dancy, G., Marchesi, B. E., Olsen, T. D., Payne, L. A., Reiter, A. G., and Sikkink, K. (2019). *Behind Bars and Bargains: New Findings on Transitional Justice in Emerging Democracies*. *International Studies Quarterly*, 63(1):99–110.

Dario Garcia-Gasulla
dario.garcia@bsc.es

Atia Cortés
Sergio Alvarez-Napagao
Ulises Cortés

