

# How biased are our perceptions of Data Policies?

Customer trust vs. GDPR compliance analysis

Group 3

Romain Besombes | Tommy Tran | Etienne Windels  
Thomas de Mareuil | Paul-Emile Dugnat



HEC  
PARIS



# WANTED

Customer Trust

&

Data Privacy

&

Compliance

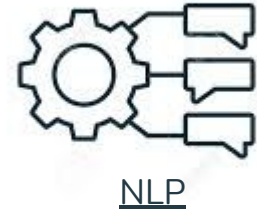
# Reward

Benefits

# What terms/expressions in a data policy makes it perceived as non compliant by the customers?

→ Customer grades in Transparency and Control vs. content of data policies.

We used **NLP** (regression using **Bag of Words tokens**) to determine what groups of words best explain low and high grades in Transparency and Control questions.



```
Feature: transparency_mean
```

```
---
```

```
[bottom tokens]
```

third party advertising networks	: -0.1352	8 occurrences
to any third party	: -0.0908	11 occurrences
if at any time	: -0.0752	11 occurrences
at any time you	: -0.0718	13 occurrences

```
Feature: I understand the processing activities the company will  
perform on my data _mean
```

```
---
```

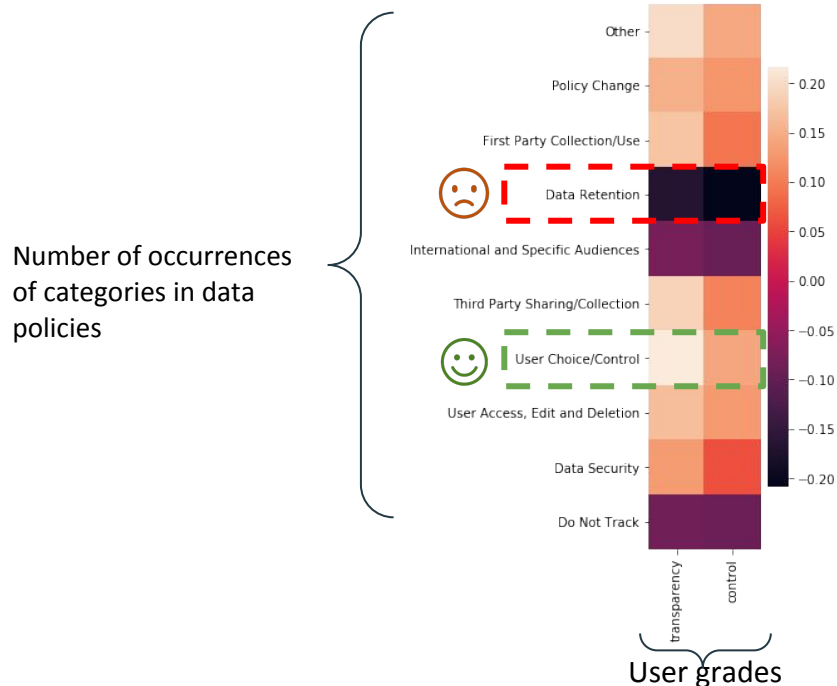
```
[top tokens]
```

privacy policy how	: 0.0276	6 occurrences
europaen union and	: 0.0264	4 occurrences
us safe harbor	: 0.0264	4 occurrences
safe harbor privacy	: 0.0259	11 occurrences

- Groups of words related to third-party data sharing impact Transparency grades negatively

# What 'categories' in a data policy makes it perceived as non compliant by the customers?

→ Correlation of Customer grades in Transparency and Control vs. 'categories' in data policies



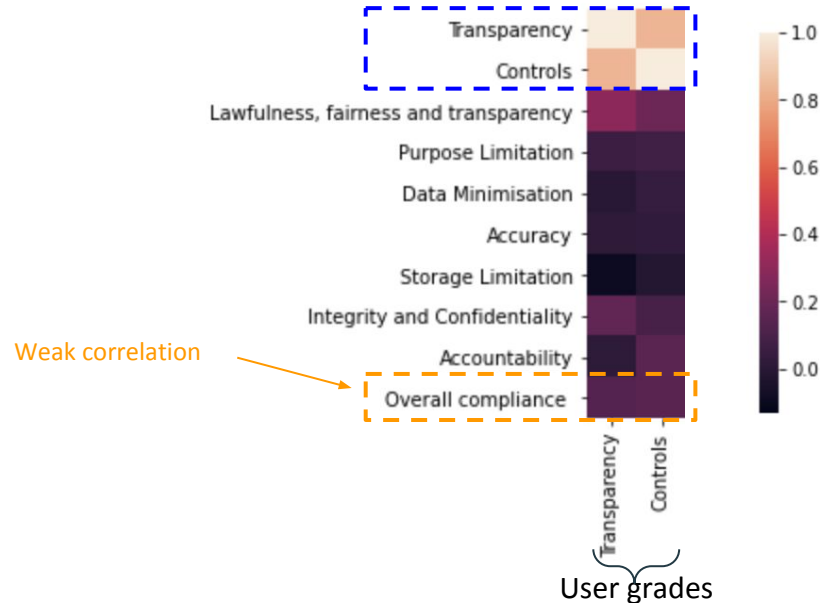
- Some topics (categories) probably frighten the users when they are included in the policy
- However, the analysis should be nuanced and refined because not seeing a category in the policy is not always a good sign

# Is customer perception of compliance correct compared to expert ratings?

→ *Customer grades in Transparency and Control vs. expert views on compliance.*

Overall compliance (experts) weakly correlated to T&C (customers). Why?

- Non-experts **don't know how to rate compliance** - ex: some technical criterias ("Storage") aren't correlated.
- Users blinded by the benefits (other / **latent variables**).



# Exploring the relationship between customer trust and other variables

- Is customer trust driven by perception of Transparency and Control?
- Or other variables come into play (perceived Benefits, variable interactions)?

We used **linear regression** to empirically study the impact of the variables on customer Trust:



Our assumptions are limited since the relationship may be non-linear and we have “little data”

	Raw	Raw + Benefits	Interaction
Features used	T + C	B + T + C	$B + T + C + B*T + B*C + C*T + C*T*B$
R square	74.1%	77.4%	77.7%
Benefit coefficient	NA	0.32	0.20
Control coefficient	0.66	0.48	-0.14
Transparency coefficient	0.51	0.47	0.52



- Perceived Transparency and Control, in addition to being biased vs. expert views, **aren't enough to explain customer Trust.**
- Perceived **Benefits** weigh in, and twist our perception... We could even lower our privacy standards if we deem benefits worth enough!

Thanks!

Questions?







# Appendices

## NLP outputs - Other examples

Feature: Integrity and Confidentiality

---

[top tokens]

any personal information	:	0.0193		40 occurrences
personally identifiable information	:	0.0188		60 occurrences
the privacy policy	:	0.0156		38 occurrences
the privacy practices	:	0.0149		30 occurrences
the security of	:	0.0143		37 occurrences
to the privacy	:	0.0143		18 occurrences
any personally identifiable	:	0.0143		23 occurrences
the information we	:	0.0139		51 occurrences
we may use	:	0.0135		54 occurrences
the personally identifiable	:	0.0132		16 occurrences

Feature: The privacy policy is clear, and I understand my privacy rights\_mean

---

[bottom tokens]

third party advertising networks	:	-0.2000		8 occurrences
to any third party	:	-0.0916		11 occurrences
at any time you	:	-0.0757		13 occurrences
if at any time	:	-0.0745		11 occurrences
to opt out of	:	-0.0275		51 occurrences
the bottom of any	:	-0.0247		4 occurrences
to third parties we	:	-0.0224		4 occurrences
the privacy of our	:	-0.0222		6 occurrences
at the bottom of	:	-0.0182		19 occurrences
you visit our web	:	-0.0174		6 occurrences