Terms of Service Text Evaluator (ToSTE)

TERMS OF SERVICE
"I HAVE READ AND AGREE TO THE TERMS". DO YOU ACCEPT?

Well yes, but actually no

It is a web extension that uses **machine learning** to analyze complicated privacy clauses in **ToS** and presents key points in simpler language to users.

CORNELL TECH | HOME OF THE JACOBS TECHNION-CORNELL INSTITUTE

Top 3 Risks and Experiments

⚠️ Key Risks	✅ Risk Mitigating Task	🧪 Experiment Type	📖 Experiment Description	⏰ Time
Do users want our product?	Run a paper prototype experiment to gauge potential users on their interest with ToS.	**Paper Prototype**	Provide users with both sample ToS and output of ToSTE, check reaction	**2 days**
Company might block our product	More research: check if we can read ToS is image vs text	**Small-world pilot**	Run a dummy, custom script that can read the text off of a website, detect if the dummy script is blocked.	**1 week**
Is B2B model feasible?	Find out if startups are willing to pay for ToSTE instead of reading the whole ToS and spend more resources on it	**Equivalent experiment**	Run an experiment using a free book and a comprehensive note, but the users pay for it.	**2 days**

Paper ToSTE

Paper prototype - Will People Use ToSTE?

RESULTS
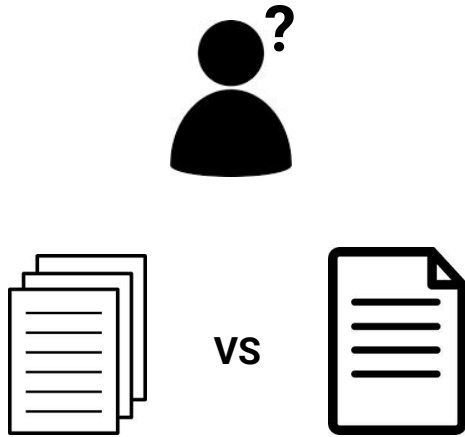
Prefer simple ToS: 82.7%
Read both: 4.3%
Not reading either one: 13%
Only read long ToS
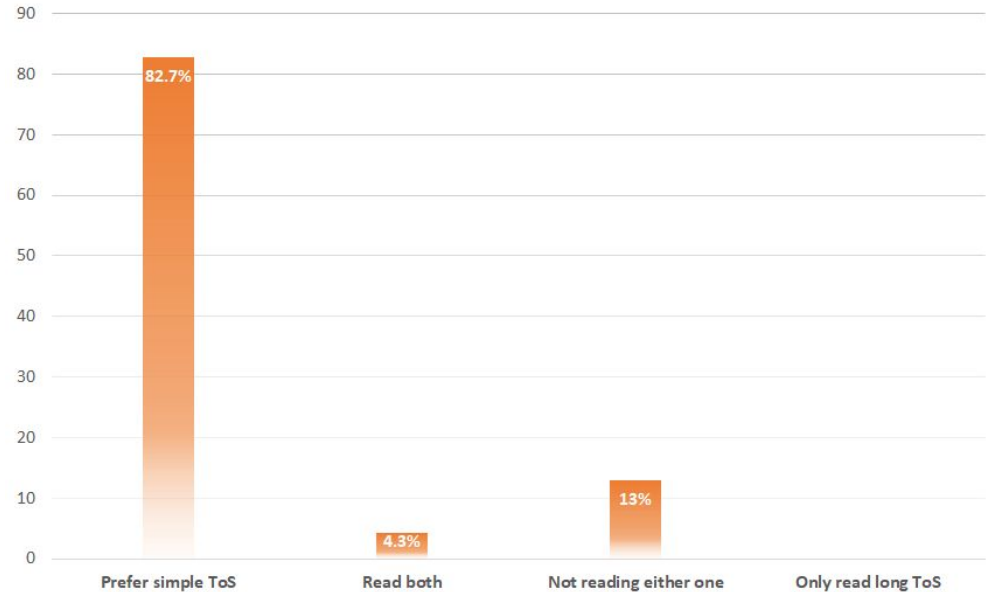
Dummy Script

Small-Word Prototype - can we access the data to build ToSTE?
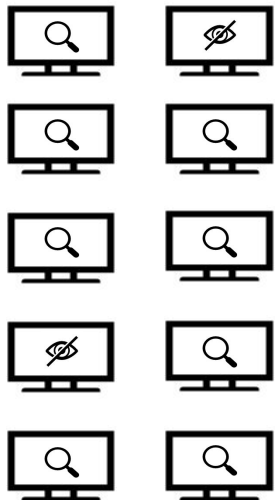
Python Dummy Script — **Success!**

JS Browser Extension — **Success!**

Main Discoveries

- Python (as a stand-alone script) & JavaScript can read static web-pages without much concern

- Pages with generally longer word counts = privacy

- No common structure for navigation amongst websites
  - Some have ToS, Privacy, and Cookies separate, some have them together - no standardized set of categories of ToS's

Sites Looked at:

Facebook, Google, YouTube ,Instagram, Twitter, Amazon

Netflix, New York Times, Spotify, Cornell Tech

Dummy Script: Results (pt.2)
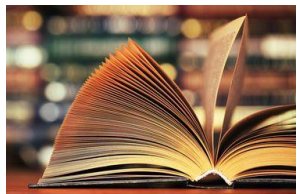
What are some remaining concerns or insights?

Average Word Counts on top 3 categories of ToS

	Python	JavaScript
Privacy	4069	4015
General ToS	3985	3929
Data	3350	4346

Book Experiment
Equivalent Experiment - Is B2B model feasible?

Free Book

Note with a price
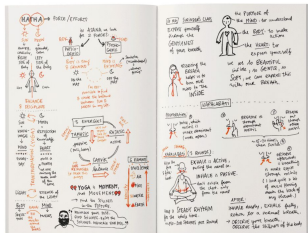
Startup 1

Startup 2

Startup 3

- Read **none of them** if the content is about terms of service.

- Read **none of them** if the content is about terms of service.

- Interested in **generating ToS** instead of reading it

- Read **none of them** if the content is about terms of service.

- Only trust law firm on generating and reviewing ToS

- Interested in summarizing ToS for users

New Idea: Can ToSTE help to improve conversion rate, encouraging people to sign up for the service?

Appendix

Python results

PYTHON	Conditions	EU-US-Swiss	Privacy	Cookies	Copywrite	Takedown	Data	Platform
Amazon	4131	1078	3416					
Google	2049		4316					
NYTimes	3914		5286	694				
Youtube	4436				598	526		
Cornell Tech	399		1511					
Instagram	3198						2252	2257
Spotify	7655		4834		410			
Facebook	4182			1799			4447	
Netflix	4144		4044					
Twitter	5742		5073	1812				
AVERAGES	3985	1078	4069	1435	504	526	3350	2257

CORNELL TECH | HOME OF THE JACOBS TECHNION-CORNELL INSTITUTE

Appendix

Javascript results

JAVASCRIPT	Conditions	EU-US-Swiss	Privacy	Cookies	Copywrite	Takedown	Data	Platform
Amazon	4102	1091	3353					
Google	2023		4302					
NYTimes	3872		5182	687				
Youtube	4420				600	531		
Cornell Tech	395		1402					
Instagram	3172						4286	2231
Spotify	7592		4806		408			
Facebook	4163			1792			4405	
Netflix	4110		4005					
Twitter	5441		5052	1783				
AVERAGES	3929	1091	4015	1421	504	531	4346	2231

CORNELL TECH | HOME OF THE JACOBS TECHNION-CORNELL INSTITUTE