Lead PI and contact: Shomir Wilson
(**shomir@psu.edu**)

Also PIs:

Lee Giles (clg20@psu.edu)

Florian Schaub (fschaub@umich.edu)

Gabriela Zanfir-Fortuna
  (gzanfir-fortuna@fpf.org)

PrivaSeer

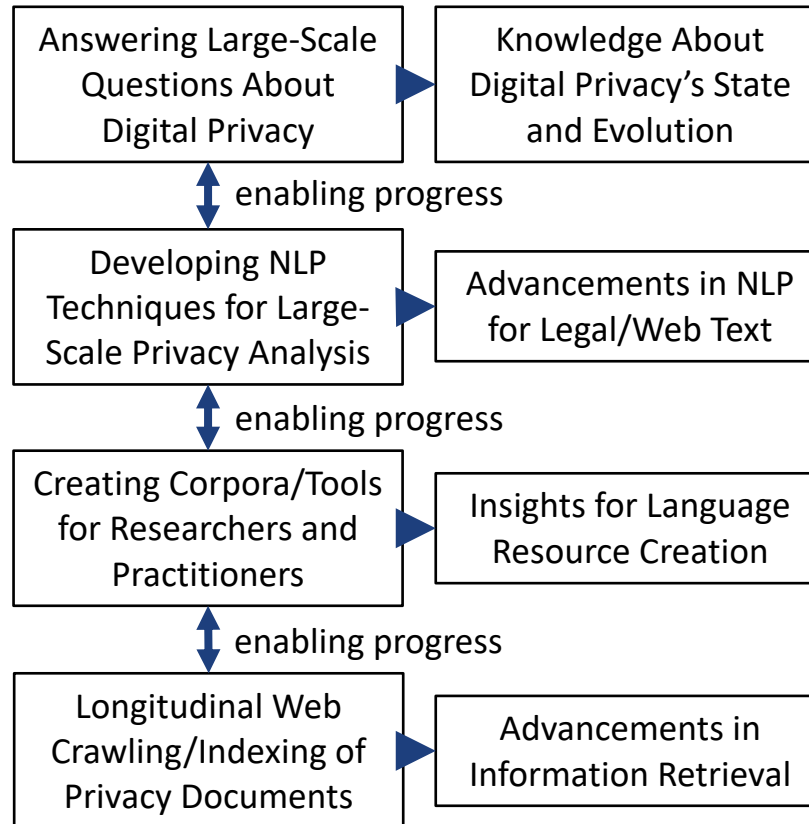## https://privaseer.ist.psu.edu/

Challenge:

A huge volume of privacy-related text exists: privacy policies, terms of use documents, cookie policies, privacy laws, and others.

Collecting and analyzing all this text is an obstacle to privacy research, practice, and effective regulation.

Solution:

We're building a large-scale, annotated, and searchable resource of privacy-related texts, while also advancing techniques for automatically interpreting them.

Answering Large-Scale Questions About Digital Privacy	▶	Knowledge About Digital Privacy's State and Evolution

↕ enabling progress

Developing NLP Techniques for Large-Scale Privacy Analysis	▶	Advancements in NLP for Legal/Web Text

↕ enabling progress

Creating Corpora/Tools for Researchers and Practitioners	▶	Insights for Language Resource Creation

↕ enabling progress

Longitudinal Web Crawling/Indexing of Privacy Documents	▶	Advancements in Information Retrieval

Scientific Impact:

Project results will enable analyses of the state of privacy at an unprecedented scale and will remove barriers for creating privacy analysis and support tools for researchers, practitioners, and policymakers.

Broader Impact and Broader Participation:

This project supports making online privacy manageable for consumers, regulators, and others invested in our information society.

Project members are recruiting and working with a diverse team of students and fellows to participate in research activities.

PennState
College of Information
Sciences and Technology

M | SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

FUTURE OF PRIVACY FORUM