The Chief Justice Meir Shamgar
Center for Digital Law & Innovation
Tel Aviv University

DESIGNING AI FOR DEMOCRACY
COMBATING DISINFORMATION BY TECHNOLOGICAL MEASURES

British Embassy Israel

Microsoft

The Chief Justice Meir Shamgar Center for Digital Law & Innovation Tel Aviv University

iCRC Blavatnik Interdisciplinary Cyber Research Center

Microsoft for Startups

bright data

מרכז אדמונד י' ספרא לאתיקה
The Edmond J. Safra Center for Ethics

איגוד האינטרנט הישראלי ISOC-IL .il

TAD The AI and Data Science Center Tel Aviv University

MEITAR | LAW OFFICES

The Buchmann Faculty of Law Tel Aviv University

About The Challenge

2024 is expected to be an election year around the world, in the course of which elections will be held in more than seventy countries with a population of circa 4.2 billion people. Against this background, the World Economic Forum recently declared that the dissemination of disinformation and misinformation constitutes the biggest threat to the stability of democracies around the world in the short term. This challenge also tops the agenda for tech companies around the world, and a cooperative venture was recently announced—the Tech Accord to Combat Deceptive Use of AI—which includes leading tech companies. In parallel, against the background of the ongoing war in Gaza and the erosion of Israel's legitimacy in the world, it is important that we develop in Israel the technological and regulatory means for contending with attacks of disinformation.

A coalition of partners, including the Shamgar Center for Digital Law and Innovation at Tel Aviv University, Microsoft Israel, Bright Data, the British Embassy in Israel, the Edmond J. Safra Center for Ethics, the Center for Artificial Intelligence and Data Science, the Blavatnik Interdisciplinary Cyber Research Center at Tel Aviv University, the Israel Internet Association, and Meitar Law Firm, led the Challenge "Designing AI for Democracy."

The goal of the Challenge is to develop technological solutions based on a legal, social, and institutional understanding of how disinformation is spread. The aim is to create tools that will assist in identifying and monitoring bots, detecting inauthentic campaigns on social media, and efficiently implementing legislative and regulatory frameworks.

Our Partners

British Embassy Israel

Microsoft

מרכז הנשיא מאיר שמגר
למשפט דיגיטלי וחדשנות
אוניברסיטת תל אביב

iCRC
Blavatnik Interdisciplinary
Cyber Research Center

TEL AVIV UNIVERSITY אוניברסיטת תל-אביב
מרכז אדמונד י' ספרא לאתיקה
The Edmond J. Safra Center for Ethics

bright data

Microsoft for Startups

MEITAR
MEITAR | LAW OFFICES | מיתר | עורכי דין

איגוד
האינטרנט
הישראלי .il
ISOC-IL

TAD
The AI and Data
Science Center
Tel Aviv University

The Buchmann
Faculty of Law
Tel Aviv
University

Advisory Committee

Adv. Hila Hubsch, Head of Corporate, External Legal Affairs, Microsoft Israel

Adv. Dana Mazia, General Manager, The Bright Initiative by Bright Data

Prof. Niva Elkin – Koren, Academic Director, The Chief Justice Meir Shamgar Center for Digital Law and Innovation, The Buchmann Faculty of Law, Tel Aviv University

Dr. Shimon Shahar, Scientific Director, The Center for artificial intelligence and Data science, Tel Aviv University

Prof. Issachar Rosen-Zvi, Head of the Edmond J. Safra Center for Ethic, Tel Aviv University

Dr. Assaf Wiener, Managing Director of Research, Policy, and Government Affairs, Israel Internet Association

Adv. Ignacio Gonzalez Royo, Partner, Meitar Law Offices

Adv. Shahd Asfour, MENA Tech Policy Lead, British Embassy

Judges

Ms. Hila Rom, Founder & Managing Partner, RUNI Ventures

Ms. Yasmin Lukatz, Founder & Executive Director, ICON; Founder and Chair of CODE FOR ISRARL

Ms. Sarit Firon, Managing Partner, Team8

Ms. Esther Barak Landes, Chair of Wingsil.com

XPOZ

Companies, communities and governments face a growing challenge to defend themselves from narrative attacks. XPOZ is an advanced AI cyber defense platform purpose-built to combat those attacks by detecting and countering disinformation at its source. The feature we present at this event focuses on identifying coordinated influence groups on social networks. Currently, the main challenge in identifying these groups stems from the wide variety of content types, the number of users, and the changing dynamics of their activities. Our solution combines AI-based algorithms that enable the detection of groups with suspicious characteristics of inauthentic activity. Once a group is identified, an in-depth analysis is conducted, including the percentage of bots in the group, the country of origin, and other essential data that provide a comprehensive picture of the nature of the activity.

Team members: Atalia Baumer, Mohar Bar, Tal Bouhnik, Ido Avnir, Ran Farhi, Erez Koren.

Humane AI

ExplainIt by Humane AI ensures that AI-driven recommendation systems respect democratic values by providing transparency, control, and the right to explanation. It allows users to understand how decisions are made by AI, granting them greater autonomy and mitigating risks such as bias and manipulation. By building a bridge between users, organizations, and regulators, ExplainIt facilitates informed, responsible decision-making while ensuring compliance with AI regulations. This product empowers both individuals and companies to engage in AI systems with clarity, trust, and democratic oversight.

Team members: Dr. Limor Ziv, Tomer Yatzkan, Eyal Hurwitz.

ECHO-VERITAS

EcoVeritas utilizes multiple AI agents to mimic the human storytelling mind, creating a sophisticated framework for textual analysis. Our system excels by combining natural language processing, narratology—the study of narrative structures—and poetic pattern recognition, setting us apart in the industry. Founded in response to the events of October 7th, EcoVeritas meets the critical need for accurate information dissemination. Our technology has proven highly effective in detecting fake news during blind tests in multiple datasets, automatically processing thousands of items per batch with a validated success rate of 98.4% to 98.9%. We examine narratives from news articles to social media content on platforms like TikTok in multiple languages. EcoVeritas offers detailed insights through services such as author integrity estimation, narrative coherence analysis, AI-authorship detection, and the identification of defamation or concealed violent content. Our upcoming "story vector" technology will further advance this analysis, providing applications across various fields—from journalism to legal settings—where verifying text integrity is essential.

Team members: Dr. Yiftach Ashkenazi, Adam Segal, Ariel Rosenblum, Na'aman Hirshfeld.

Relaible.AI

We are four experienced data scientists, addressing media bias. We developed an AI system that analyzes news articles, quantifies they express towards a particular subject, and shows how that stance is affected, by topic coverage, framing and linguistic and writing tactics (e.g., the use of sensational headlines). Our PoC, focusing on international coverage of the Iron Swords war, demonstrates the system effectiveness. We aim to use our solution to empower both journalists and the public, ultimately strengthening democratic discourse through more balanced and transparent news reporting.

Team members: Yael Gesser, Sergei Merson, Gal Ben Ari, Shahar Cohen.

VerifAI

VerifAI by Dtect is an advanced software solution aimed at enhancing the reliability of digital information through sophisticated AI-driven text analysis. Our technology specializes in automated fact-checking and real-time verification of digital content to detect and alert against misinformation and disinformation. Led by Dor Shkedi, who brings leadership skills from his military background in a special unit in the Navy and experience in entrepreneurship, and Roee Dar, an expert in artificial intelligence with a background in elite military technology units and startups.

Team members: Roee Dar, Dor Shkedi, Lior Moyal.

InfluenceBlocker

InfluenceBlocker addresses the complex challenge of foreign interference and disinformation campaigns that exploit private (one-on-one) communication. Through monitoring, identification, and characterization methods based on activity signals indicating private communication, InfluenceBlocker tackles the challenge of identifying and tracking these operations, which cannot be uncovered through public data. The solution leverages the team's extensive experience gained through years of civic, academic, volunteer, and professional research into malicious online activities and democratic processes. While prioritizing privacy and ethics, the initiative not only protects public discourse from foreign actors seeking to covertly influence and undermine democracy but also safeguards innocent citizens who are targeted.

Team members: Nitzan Yasur, Inbar Yasur.

Rhetoric AI

Rhetoric AI is building new methods for automatically evaluating and generating persuasive arguments. We augment language models with strategic planning and reasoning capabilities (as in OpenAI's most recent o1 model) by extending the Tree-of-Thoughts method from Yao et al., 2023. Our research is supported by the Technion, an ERC grant, and industry partners Wand AI and DebaterHub. Currently, we are conducting a randomized controlled study comparing the quality and persuasiveness of our AI-generated arguments against those of top European and World Championship debaters. Committed to advancing the NLP research community, we openly share our methodologies and tools, available on GitHub (https://github.com/zbambergerNLP/strategic-debate-tot). Our mission is to develop persuasive argumentation technologies which can advance marketing efforts, assist lawyers, encourage healthy practices, and combat anti-democratic rhetoric at scale.

Team members: Zachary Bamberger, Yosef Ben Yehuda, Ze'ev Sheleg, Dmitry Rudman.