

Looking Back On 20 Years Of Data Mining The Web And Forward To Our AI-Powered Online Future

Dr. Kalev Hannes Leetaru

Founder, GDEL Project

Media Fellow, Real Clear Foundation

Senior Fellow, Center for Cyber & Homeland Security, The George Washington University

Google Developer Expert for Google Cloud Platform

What happens when massive computing power brings together an ever-growing cross-section of the world's information in realtime, from news media to social media, books to academic literature, the world's libraries to the web itself, mass machine translates all of it from more than 100 languages and transforms this immense record of humanity into a living global catalog of our planet, connecting the world's information into a single massive ever-evolving realtime network that allows us to peer into the very soul of global society? Today the GDEL Project (<https://www.gdelproject.org/>) is one of the largest open datasets for understanding human society, totaling more than 3.2 trillion datapoints spanning 200 years and has become a global standard used by humanitarians, NGOs, scholars, journalists and even ordinary citizens to make sense of our chaotic and rapidly evolving world. From disaster response to countering wildlife crime, epidemic early warning to food security, estimating realtime global risk to mapping the global flow of ideas and narratives, GDEL explores how we can use data to form bridges that can help build empathy and expand our own limited horizons, breaking down linguistic, geographic and cultural barriers to let us see the world through the eyes of others and even forecast the future, capturing the realtime heartbeat of the planet we call home. From mining thousands of web pages on a single small server 23 years ago to exploring our humanity through trillions of datapoints spanning data centers in 12 countries today, we'll take a journey through what its been like to conduct web-scale research over the past two decades, from the days when Mosaic ruled the web to today's globalized cloud and what we've learned from all those studies about what makes us human. Along the way we'll look at how traditional machine learning and statistical models transforming billions of news articles into hundreds of millions of human events, tens of billions of hyperlinks and trillions of knowledge graph entries have been joined by deep learning approaches capable of translating half a billion images totaling a quarter-trillion pixels into 300 billion datapoints recording the objects, activities, locations, words and emotions through which we see the world around us. The ability of the emerging world of deep learning to lend structure to content that has never before been computationally explorable, on through systems capable of asking questions of our data and understanding its deeper patterns entirely on their own, we are reaching a world in which the web is increasingly becoming accessible in ways we couldn't dream even a few years ago. Here's what it looks like to conduct data analytics at a truly planetary scale and the incredible new insights we gain about the daily heartbeat of our global world and how our AI powered online future will help us make sense of our world in ways we could never have imagined.

ABOUT THE SPEAKER



Dr. Kalev Hannes Leetaru - One of Foreign Policy Magazine's Top 100 Global Thinkers of 2013, Kalev is a Media Fellow at the Real Clear Foundation, a Senior Fellow at the George Washington University Center for Cyber & Homeland Security and a member of its Counterterrorism and Intelligence Task Force and a Google Developer Expert for Google Cloud Platform. From 2013-2014 he was the Yahoo! Fellow in Residence of International Values, Communications Technology & the Global Internet at Georgetown University's Edmund A. Walsh School of Foreign Service, where he was also an Adjunct Assistant Professor, as well as a Council Member of the World Economic Forum's Global Agenda Council on the Future of Government. His work has been profiled in Nature, the New York Times, The Economist, BBC, Discovery Channel and the presses of more than 100 nations. In 2011 The Economist selected his Culturomics 2.0 study as one of just five science discoveries deemed the most significant developments of 2011. Kalev's work focuses on how innovative applications of the world's largest datasets, computing platforms, algorithms and mind-sets can reimagine the way we understand and interact with our global world. More on his latest projects can be found on his website at <https://www.kalevleetaru.com/> or <https://blog.gdeltproject.org>