

# Let's Teach AI to be Kind

EthicsNet Guardians' Challenge  
By Bryn McCoy, Founder Lux Lab

We're delighted to join you in this challenge, and to work with so many brilliant leaders, all working to nurture the greater good.

"We have to work to make artificial intelligence not just intelligent, but focused on the right data sets to be ethical and good."

–Mo Gawdat



To inform our design direction, we kicked off a research study to identify people's thoughts, feelings and experiences with kindness.

The data collected helped us identify + define the biggest opportunities to explore.

We worked to understand–

Why are we teaching kindness?

How does kindness impact a person; how does it impact a community?

What triggers kindness?

## Primary research–

Survey 'how have you taught, learned or experienced kindness'

Interviews with life coaches on teaching kindness

## Secondary research–

Media trend analysis for sharing kindness

## Why do we teach kindness?

Teaching kindness often starts to share principles of social good, then continues as a conversation, both interpersonal and as central themes in cultural media, to help others and uphold fairness.

## How does kindness impact a community?

At a surface level, kindness inspires reciprocity in friendship and warmth. Many people follow these basic rules of social norms to achieve acceptance in our community. Deeper motivations about kindness lead to loving generosity which develops trust and benevolence.



## What triggers kindness?

Social expectations, good-natured personality, altruism and compassion spark kindness.

Acts of kindness generate gratitude. Gratitude inspires action to deliver more kindness, either in reciprocity or through acts of benevolence.

## Kindness journey

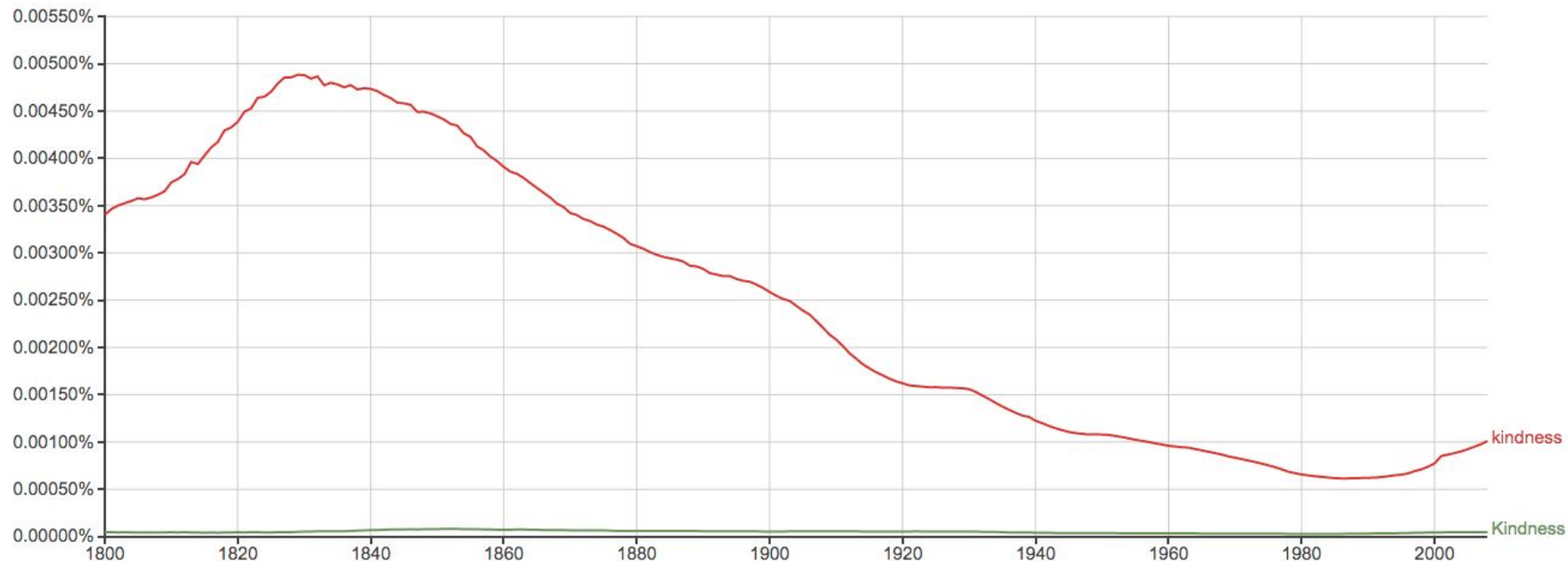
How would a person or AI interact with kindness?

The kindness experience is an iterative cycle:

Experience kindness > experience/acknowledge gratitude >  
categorize/synthesize the kindness/gratitude exchange > share/teach  
kindness

# Trending Kindness

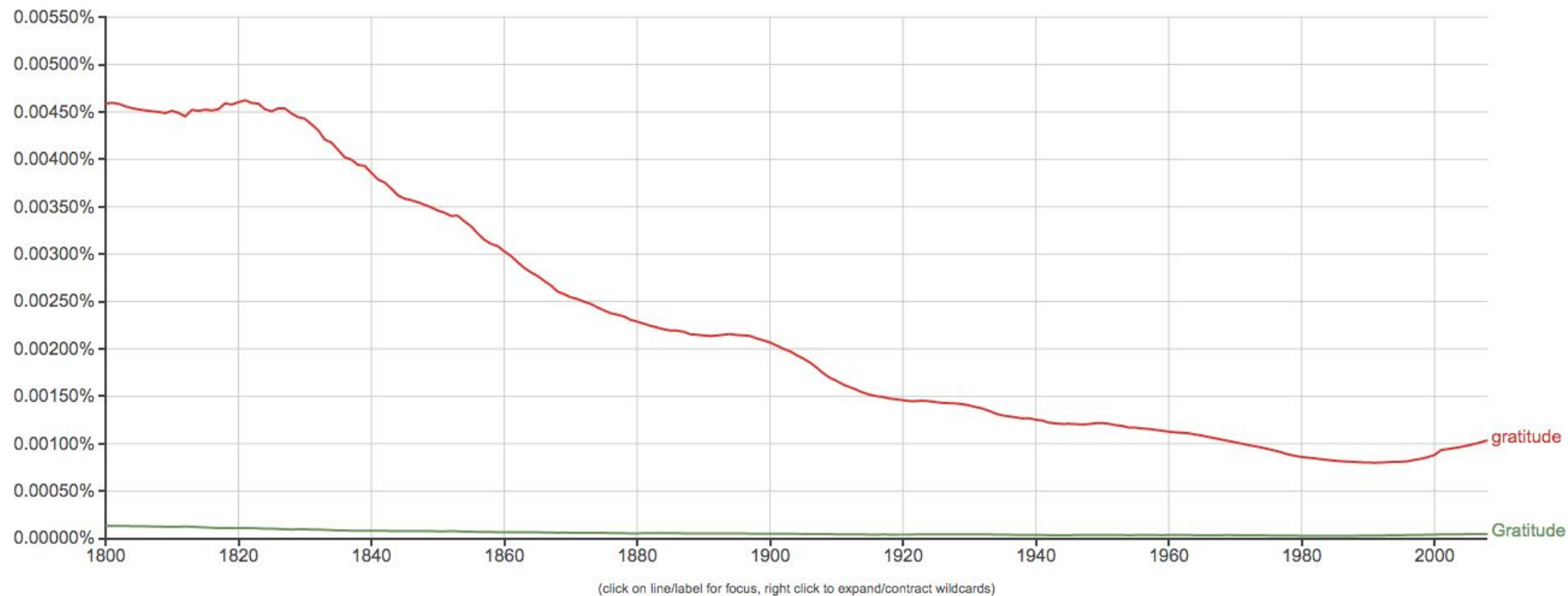
Instances of the term 'kindness' in the Google Books archive



(click on line/label for focus, right click to expand/contract wildcards)

# Trending Gratitude

Instances of the term 'gratitude' in the Google Books archive



Trends in cultural references of kindness and gratitude show a similar trajectory.

We are currently seeing a positive trend in both.

How might we design a prosocial data set?



# Data set design principles

# Identify prosocial data

Expression of behavior intended to benefit another including helping, sharing, volunteering, donating, conforming to socially accepted behaviors is prosocial.

Source: Google

Expression of empathy fosters prosocial behavior as well.



# Avoid bias in the data set

Kicking off this global conversation to encompass diversity in attitudes, perspectives and opinions is a great place to start.

Continue to be mindful about our sources for analyses, to ensure we're cultivating values and opinions from a population spanning many locations and cultures.

# Test and iterate responsibly

Take responsibility for the outcome of experiments leveraging this data set and be accountable to serving the common good.

How might we develop a large, equally distributed data set?

# Analyze cultural expressions of kindness + gratitude

Communities express forms of culture through art, music, dance, ritual, religion, and technologies such as tool usage, cooking, shelter, and clothing.

Source: Wikipedia

Analyses of the following to develop our understanding of the human values linked to cultural artifacts–

Artist statements in public collections and published critic reviews

Published mission statements and market outcomes in annual reports from leading brands of our iconic tools

Consumer reviews of these artifacts and experiences

# Trend analysis across media channels

Scrape data to build a data set across existing publications, for example:

Books – Google Books has indexed ~23.2MM books referencing 'kindness' and ~21.4MM books referencing 'gratitude'

Academic research – Google Scholar has indexed ~3,400 articles focused on kindness research

Sentiment analysis in social media streams – greater than 3,000 ongoing daily 'kindness' and 'gratitude' conversations on Twitter; ~4.9MM 'kindness' + ~13.3MM 'gratitude' posts on Instagram

# Ethnographic studies

Observation and conversation with people about kindness and gratitude in their own environments is likely to produce the most culturally diverse and complete data.

TEGA – launch a study with this network of technology-enabled girl ambassadors to research authentic community dynamics in hard-to-reach communities across Nigeria, Malawi, Rwanda, India, Bangladesh and the USA

Diary studies – invite participants to journal and scrapbook about their experiences of kindness and gratitude, follow up with in-person or video interviews

# Continue the kindness campaign

Continue to research the results of existing campaigns that seek to identify and develop kindness and gratitude. From Oprah's Thank You Game to the many personal campaigns across social media, there's no shortage of data to capture.

Launch an app with a conversational experience to invite ongoing contribution and interaction. Make it your own campaign to have a hand in prioritizing this message. All the brilliant minds in the world might be grateful.

Thank you.

Bryn McCoy

bryn@luxlab.ai

A black and white photograph of a beach. In the foreground, a white rectangular sign is placed on the sand, tilted slightly. The sign has the words "KINDNESS" and "IS MAGIC" printed in bold, black, sans-serif capital letters. The background shows a calm body of water meeting a distant shoreline with some buildings and trees. A large, dark shadow of a person is cast on the sand and water to the right of the sign.

**KINDNESS  
IS MAGIC**