

How Human Factors Can Improve Neural IR

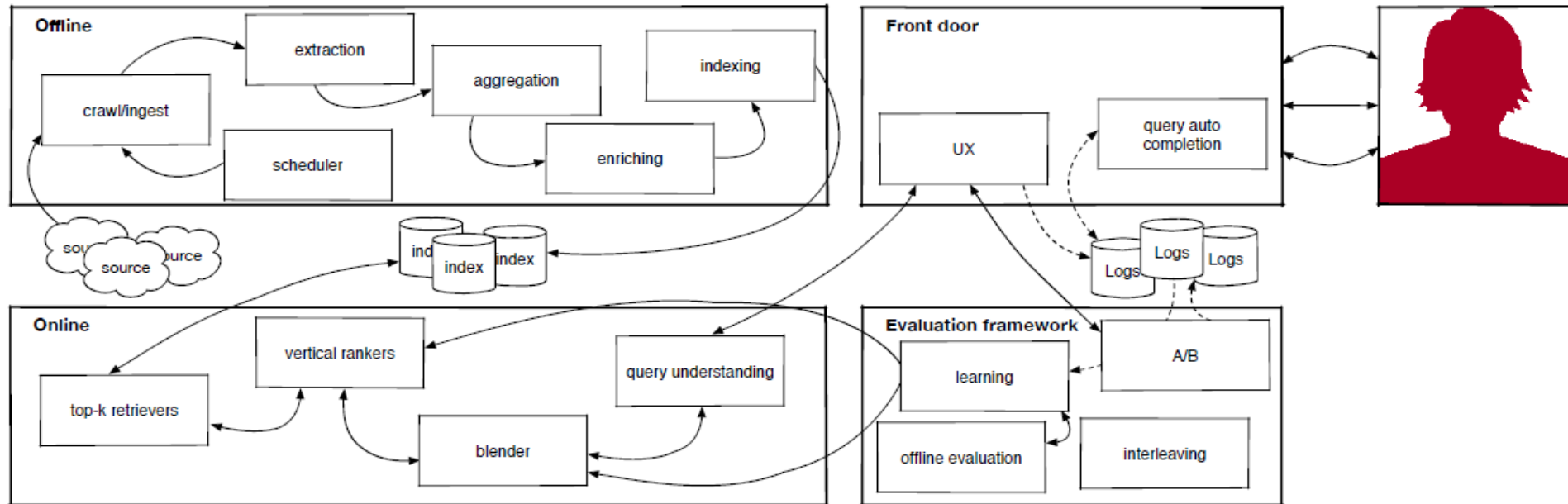
Formerly known as

IR Intelligence: Introduction to Neural IR & Learning to Rank

Search Solutions 2020



Learning to Rank Model





Michael Bendersky

FOLLOW

Google Research

Verified email at google.com - [Homepage](#)

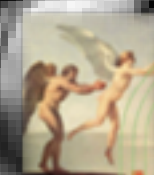
[Information Retrieval](#) [Natural Language Processing](#) [Web Search & Data Mining](#) [Machine Learning](#)

TITLE	CITED BY	YEAR
Discovering key concepts in verbose queries M Bendersky, WB Croft Proceedings of the 31st annual international ACM SIGIR conference on ...	304	200
Learning concept importance using a weighted dependence model M Bendersky, D Metzler, WB Croft Proceedings of the third ACM international conference on Web search and data ...	198	201
Quality-biased ranking of web documents M Bendersky, WB Croft, Y Diao Proceedings of the fourth ACM international conference on Web search and ...	126	201
Analysis of long queries in a large scale search log M Bendersky, WB Croft Proceedings of the 2009 workshop on Web Search Click Data, 8-14	123	200
Parameterized concept weighting in verbose queries M Bendersky, D Metzler, WB Croft Proceedings of the 34th international ACM SIGIR conference on Research and ...	122	201
Learning to rank with selection bias in personal search X Wang, M Bendersky, D Metzler, M Najork Proceedings of the 39th International ACM SIGIR conference on Research and ...	113	201



3







See It Go

"Look," said Dick.

"See it go.

See it go up."



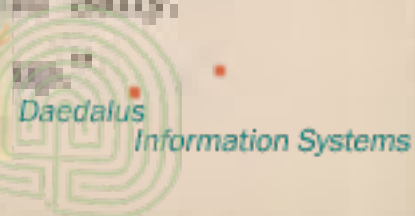
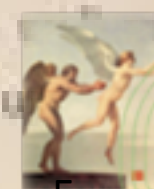
Jane said, "Oh, look!

See it go.

See it go up."

"Up," said Sally,

"Go up up."



INFORMATION RESEARCH

an international electronic journal

ISSN 1368-1613

Volume 10 No 2 January, 2005

Editorial

Papers presented at ISIC 2004: the 5th Information Seeking in Context Conference, Dublin, Ireland, 1-3 September, 2004

Keynote address: Carol Kuhlthau

Towards collaboration between information seeking research and information retrieval






Christopher Manning @chrmanning · Oct 29

Replying to @msweeny and @MDoornenbal

The status and role of consciousness is a complex one—perhaps see frontiersin.org/articles/10.33...—but more than I intended to address here; but I think we must certainly reject being able to execute some list of tasks as a sufficient criterion for intelligence



Artificial Intelligence: Does Consciousness Matter?

Artificial Intelligence: Does Consciousness Matter? Consciousness plays an important role in ...

frontiersin.org

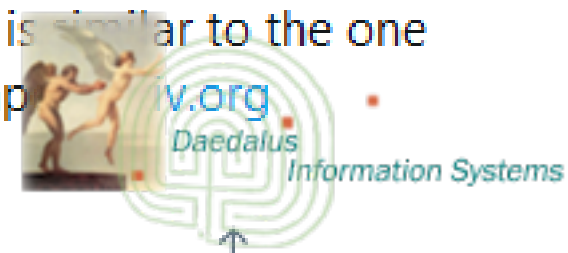


Christopher Manning @chrmanning · Oct 29



Replying to @stanfordnlp @msweeny and @MDoornenbal

As someone notes in a later comment, even though I wasn't thinking of it at the time I wrote my definitions, the position I adopt is similar to the one [@fchollet](#) argues for in much greater detail in his paper [/pdf/1911.01547...](https://arxiv.org/pdf/1911.01547...)





François Chollet

[Google, Inc.](#)

Verified email at google.com

ARTICLES

CITED BY

CO-AUTHORS

TITLE

[keras](#)

F Chollet

[Xception: Deep learning with depthwise separable convolutions](#)

F Chollet

Proceedings of the IEEE conference on computer vision and pattern ...

[Deep Learning with Python](#)

F Chollet

Manning Publications

[Tensor2tensor for neural machine translation](#)

A Vaswani, S Bengio, E Brevdo, F Chollet, AN Gomez, S Gouws, L Jones, ...
arXiv preprint arXiv:1803.07416

[Deep Learning with R](#)

F Chollet, JJ Allaire

Manning Publications

[Deepmath-deep sequence models for premise selection](#)

G Irving, C Szegedy, AA Alemi, N Eén, F Chollet, J Urban

Advances in Neural Information Processing Systems, 2235-2243

On the Measure of Intelligence

François Chollet *

Google, Inc.

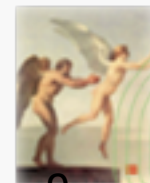
fchollet@google.com

November 5, 2019

Abstract

To make deliberate progress towards more intelligent and more human-like artificial systems, we need to be following an appropriate feedback signal: we need to be able to define and evaluate intelligence in a way that enables comparisons between two systems, as well as comparisons with humans. Over the past hundred years, there has been an abundance of attempts to define and measure intelligence, across both the fields of psychology and AI. We summarize and critically assess these definitions and evaluation approaches, while making apparent the two historical conceptions of intelligence that have implicitly guided them. We note that in practice, the contemporary AI community still gravitates towards benchmarking intelligence by comparing the *skill* exhibited by AIs and humans at specific tasks, such as board games and video games. We argue that solely measuring skill at any given task falls short of measuring intelligence, because skill is heavily modulated by prior knowledge and experience: unlimited priors or unlimited training data allow experimenters to “buy” arbitrary levels of skills for a system, in a way that masks the system’s own generalization power. We then articulate a new formal definition of intelligence based on Algorithmic Information Theory, describing intelligence as *skill-acquisition efficiency* and highlighting the concepts of *scope*, *generalization difficulty*, *priors*, and *experience*, as critical pieces to be accounted for in characterizing intelligent systems. Using this definition, we propose a set of guidelines for what a general AI benchmark should look like. Finally, we present a new benchmark closely following these guidelines, the Abstraction and Reasoning Corpus (ARC), built upon an explicit set of priors designed to be as close as possible to innate human priors. We argue that ARC can be used to measure a human-like form of general fluid intelligence and that it enables fair general intelligence comparisons

Human Intelligence



Sense-Making in our Post AlphaGo World



Sense-Making in our Post AlphaGo World



Information Behavior Mai (2016)

Quality of Information is part of a spectrum

Data >> Information >> Knowledge

Quality depends on individual characteristics

- Contextual
- Situational
- Environmental
- Emotional



Consciousness

Humans have a knowledge of core concepts through experiencing the physical world

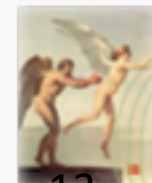
Consciousness allows for building more robust mental models that enable inference and prediction

Human consciousness entails: (that ML does not possess)

- Introspection (self awareness)
- Empathy
- Transfer learning
- Adaptation
- Novelty
- Ambiguity



Neural IR

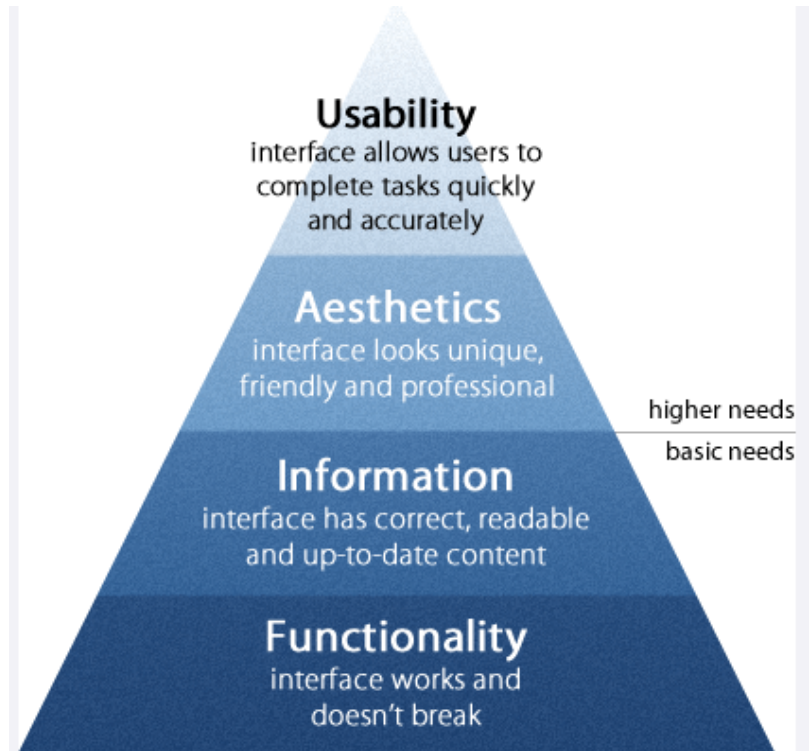
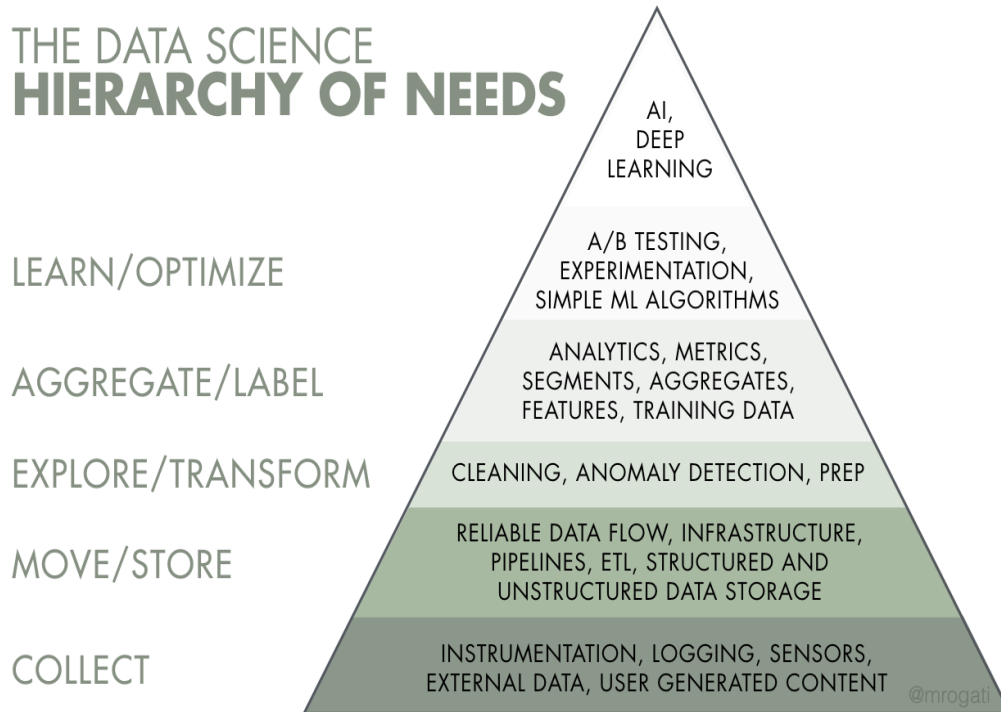


Dueling Hierarchies of Needs

AI

User Experience

THE DATA SCIENCE HIERARCHY OF NEEDS



Neural IR Intent Focus

Compositionality Principle

Scan Browsing: information scanned based on relevance to changing tasks or transient information goals (Berrypicking)

Review browsing: some information is integrated into goal after deeper review (interest)

Customer intent actions:

- Query terms & refinements
- Dwell time
- # of results considered
- Time to first action
- Click counts
- First result clicked rank



Neural IR Intent Deconstruction

Query intent = individual words that are possible indicators of customer intent. Uses term cooccurrence (proximity) models to improve retrieval relevance

Intent words = articulated by customer to refine their information needs

Content words = core topic of query

Content unites further specify the need; intent units further modify the need in one of many possible ways



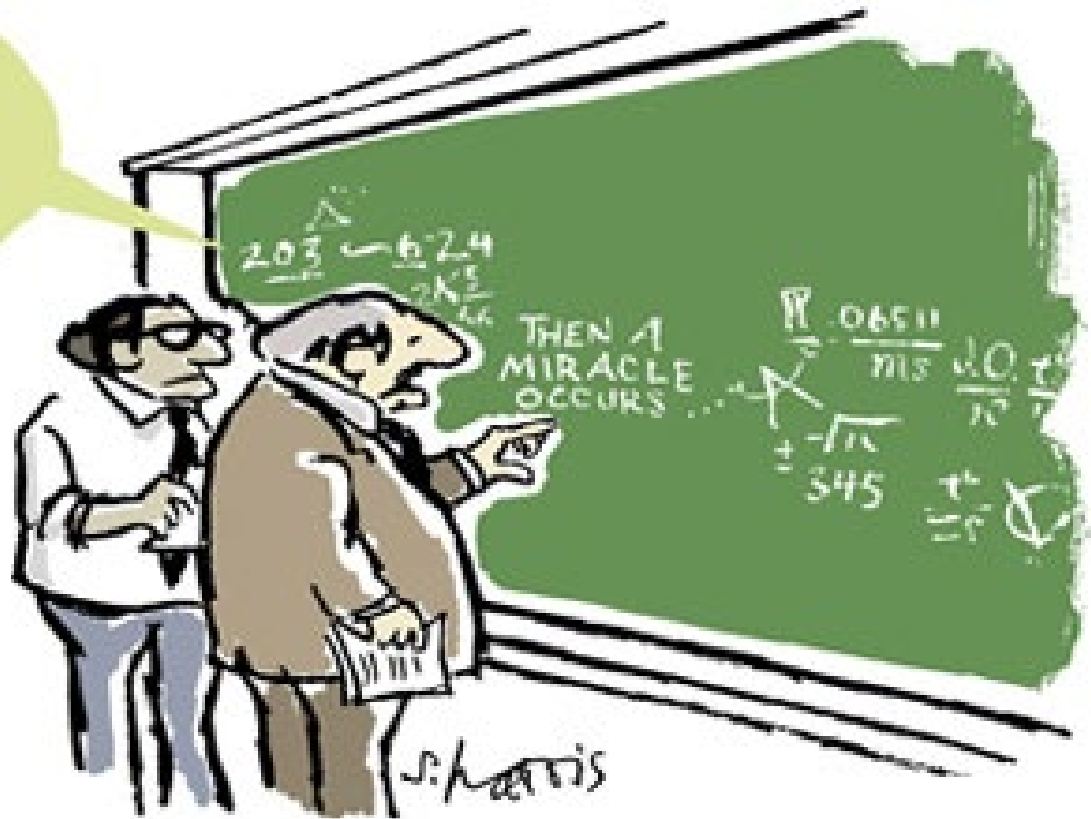
Design for Neural IR



17



I THINK YOU SHOULD BE MORE SPECIFIC HERE IN STEP TWO



Daedalus
Information Systems

Design for Machine Learning (together)

Define learning problem

- Inputs
- Outputs
- Types of training data needed

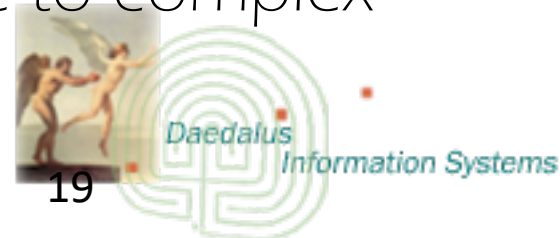
Generate good data

- Completeness
- Accurate
- Consistent
- Timely

Sketch out user and data flow (decision trees)

Test assumptions against prototype

Start with simple mechanism and move to complex



Start with Guiding Principles for AI

Principles	Applications

Principles are objectives and goals, the end state

Notes are the system rules, logic, rewards and feedback loops



Create Persona with an Empathy Profile

Find qualitative dimensions that define user understandings

Represent the emotions, not just the needs

Iterate as system learns





Philippe Hudson

Global IT Evangelist at big software enterprise

45 years, +10 years experience, loves BBQ and gadgets

Scenario

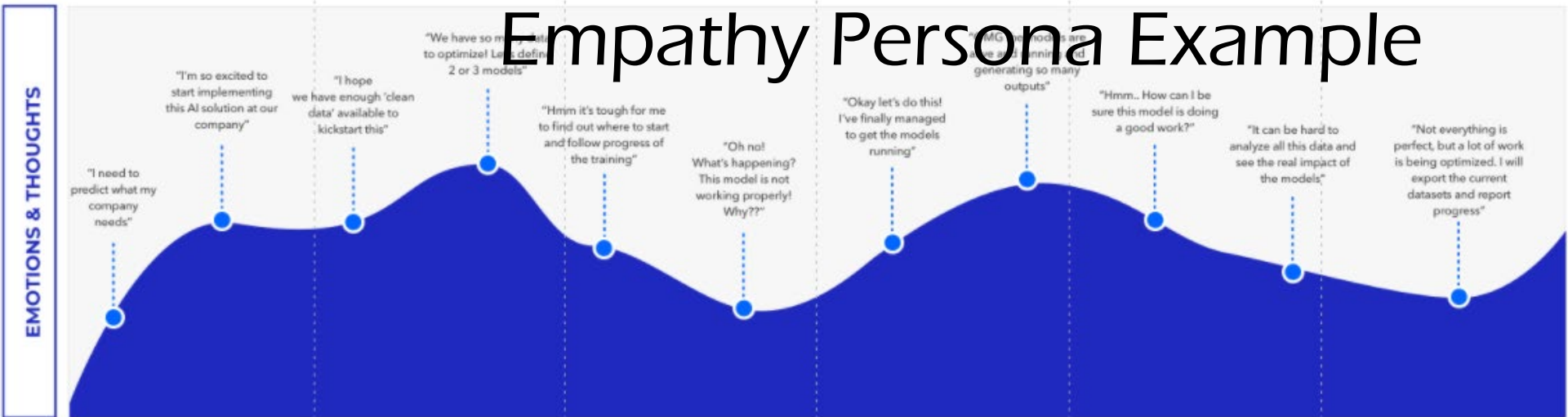
Philippe is the responsible for the AI implementation at his company. He needs to define the models that will be used, and report to the Executive Board.

Goals


- Help *knowledge workers* be more productive
- Reduce risk from error-prone processes
- Reduce company costs



Empathy Persona Example

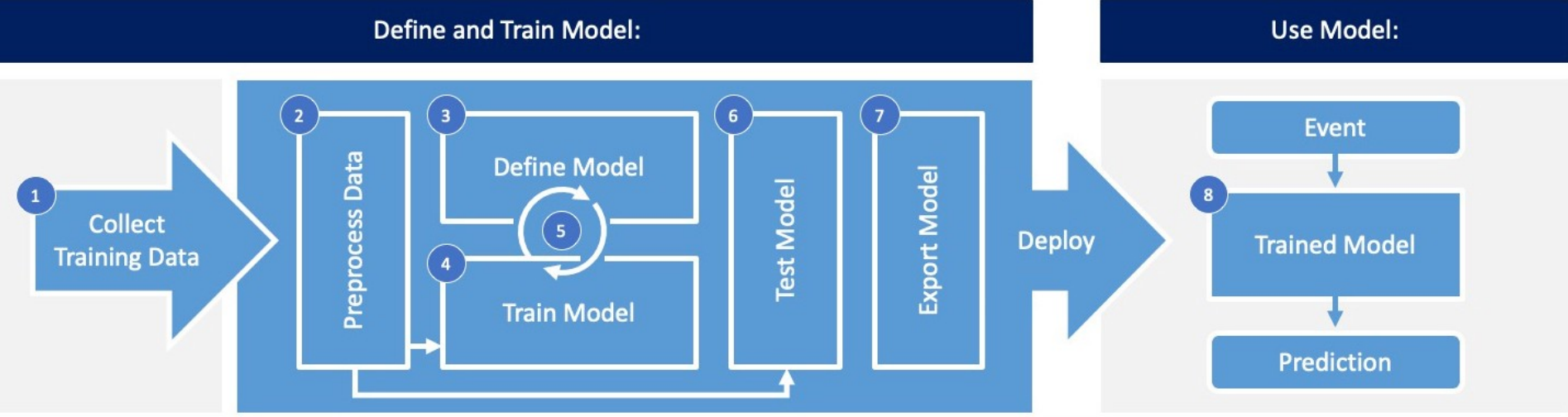


ACTIONS	USE CASE DEFINITION	MODEL DEFINITION	TRAIN MODELS	TEST MODELS	TRACK & ANALYZE	EVALUATE & REPORT
	<ul style="list-style-type: none"> • Analyse the business needs and windows of opportunity on the current data. • Understand the company goals in terms of what can be improved with Machine Learning Custom Models. 	<ul style="list-style-type: none"> • Start with a set of 20.000 documents to train and evaluate the model • Define the inputs and outputs that will be the focus of Machine Learning intervention. 	<ul style="list-style-type: none"> • Activate the model • Follow the training results • Redefine if necessary • See if the performance is evolving as expected 	<ul style="list-style-type: none"> • Analyse if the Model suggestions are being selected • Select suggestions, having more accurate and complete data quicker. • Human validation 	<ul style="list-style-type: none"> • See forms filling duration and the number of documents generated, with the help of the models • Focus on the model performance evolution • Know why the model went down 	<ul style="list-style-type: none"> • Extract data results • Export CSV file • Prepare report based on the analysis conclusions. • Present the value of the models

OPPORTUNITIES	USE CASE DEFINITION	MODEL DEFINITION	TRAIN MODELS	TEST MODELS	TRACK & ANALYZE	EVALUATE & REPORT
	<ul style="list-style-type: none"> • Everything is possible. • Although this scenario seems perfect, almost always, there are important steps that are forgotten or underestimated. • Not that many people understands yet the full and real potential of machine learning.. 	<ul style="list-style-type: none"> • On configuration, based on the dataset selected, the user may have some suggestions and alerts. • Inform the user if some fields haven't enough content to train the model with the specified requirements • Suggest catalogue more data • Need to see the amount of data filled in. 	<ul style="list-style-type: none"> • If there are errors, understand what needs to be changed. • Solve possible blockers. • How to compare the results before/after redefining some inputs/outputs? • How to compare the results before/after redefining some inputs/outputs? 	<ul style="list-style-type: none"> • If something starts to go wrong at certain moment, the 'models manager' should receive an alert to analyse and redefine certain properties of the models • When going through forms, with human filled data and ML filled data, it might be necessary to separate both content to have a clear view of what the machine filled. 	<ul style="list-style-type: none"> • Consider a dashboard for full context analysis, being able to understand the real ROI •  How can we help to prepare that report? 	<ul style="list-style-type: none"> • Information needs to be organized • A full report may be necessary to present to the board the results of this experience, showing if this investment bring a good return and costs reduction to the company.



Map out the AI Journey



Description of Workflow:

- 1** Collect training data with needed volume and quality (fit for purpose and diverse enough to avoid a biased model later in the process).
- 2** Process data in such a way that relevant features for the problem statement are identified and prepared. Also separate data into training and validation set.
- 3** Choose appropriate algorithm(s) for the given problem. Design model and tune parameters.
- 4** Model is trained based on the given data. Validate how well the model performs.
- 5** Iteratively modify the model until it performs well against the training and validation data.
- 6** Test the model against the hold out set from the collected data. Then export the model for prod
- 7**
- 8** Expose the trained model to new data / events. Model will return a prediction accordingly. If a feedback loop is included, model could also improve itself.

[Home](#) / [Talks](#) / IA at the Helm: Leading with Information

IA at the Helm: Leading with Information



[Bob Boiko](#)

IA Summit 2018 Main Conference Talk

Topic(s): career development, information architecture, and strategy



Daedalus

Information Systems

Use Soft Information Architecture

Design for evolutionary processes

- Virtual reality
- Interactive architecture
- AI (self-organizing, educating, revising technology)

Status Quo Architecture= goal oriented

Soft Architecture = behavior-based

Designer must think within, not just about the system

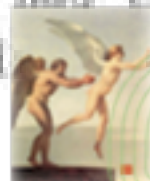


Bolo: Heuristic Traps



Name: Siegel, Benjamin
Race: White Sex: Male
DOB: 2/28/1906
Height:
Weight:
Hair: Brown
Eyes: Blue
ID No.:
LKA: Beverly Hills, CA

could have been wanted for any number of mafia related crimes from the 1940's up to and including murder. However he had the unfortunate fate of being shot to death in a Beverly Hills, CA home of his girlfriend on June 20, 1959.



Daedalus

Information Systems

Why This?
Why Now?



27





Tweet

Artificial intelligence



MIT CSAIL
@MIT_CSAIL

Algorithmic Bias

New study shows strong gender bias in image-recognition algorithms from Google, Amazon & Microsoft.

Photo of man: "official," "speaker"

Photo of woman: "hairstyle," "smile"

Paper: bit.ly/36V8kJZ

More: bit.ly/38U31x0 (v/@tsimonite @WIRED)



Tweet your reply

Search Twitter



What's happening

Television - Last night
Supernatural airing on
The CW



Trending with
Dean

#Animaniacs
All New Episodes Now Streaming
Promoted by The Animaniacs

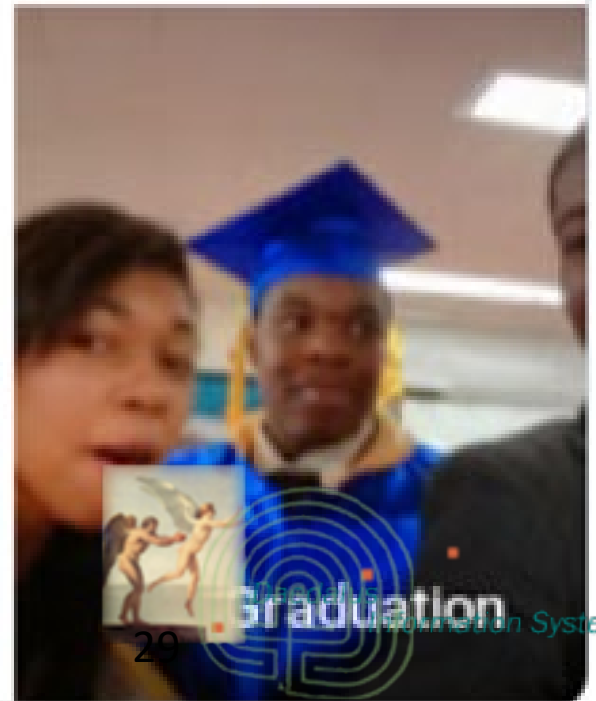
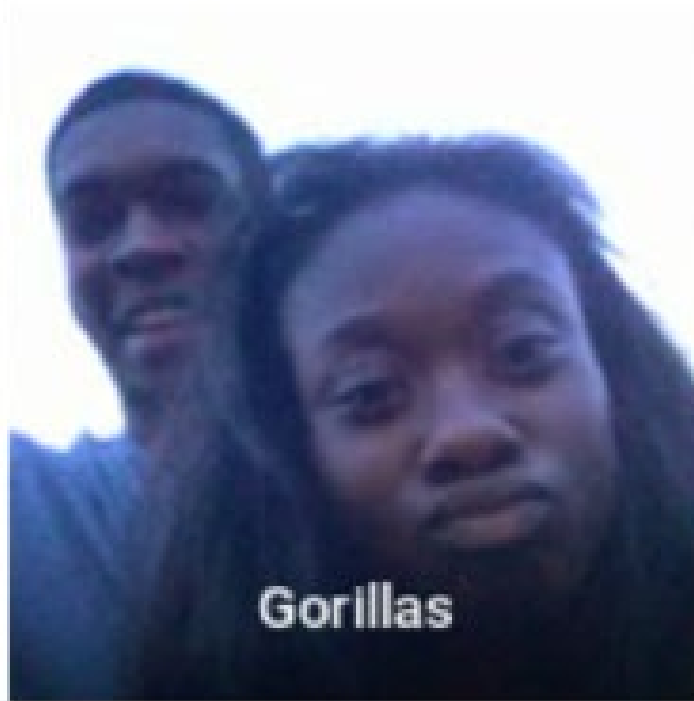
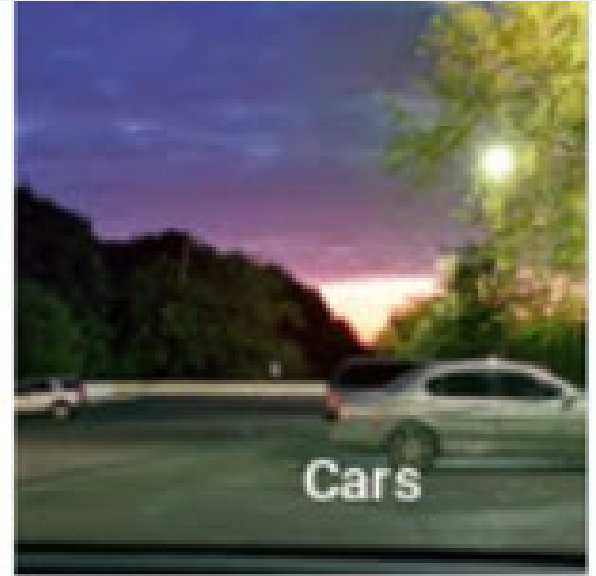
Politics - Trending
Dershowitz
6,319 Tweets

Trending in United States
Emergency Use Authorization
6,749 Tweets

US elections - LIVE
Washington: Election
news and updates



Use Cases that are too Narrow

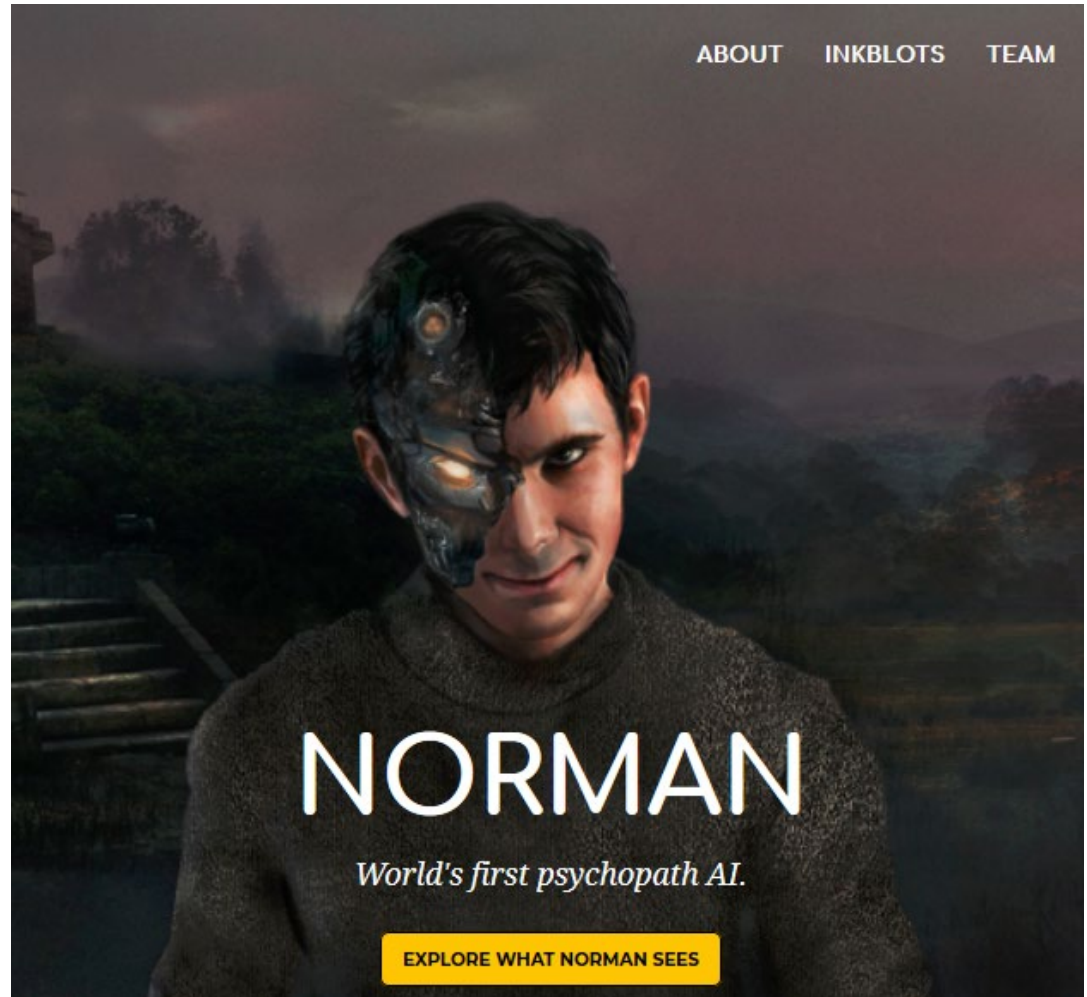


Poorly Conceived Objective Outcome

Built as a proof of concept for AI gone wrong with biased data

MIT AI Lab

Dataset was a sub-reddit dedicated to document the “disturbing reality of death.”



Poorly Designed Training Data



Bill Slawski @bill_slawski · 12s

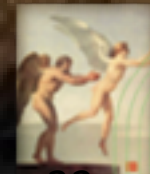
Microsoft unveils a better-behaved chatbot after its last one became a NAZI



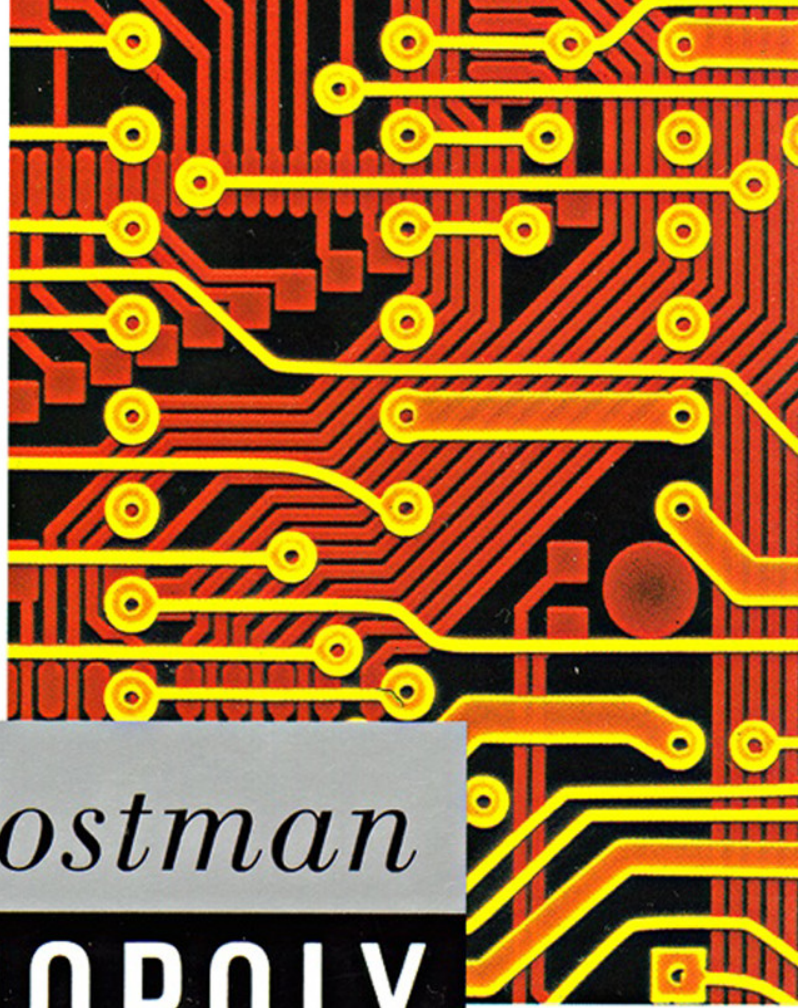
Microsoft unveils a better-behaved chatbot after its last one became ...

Tech giant takes another pop at the artificial intelligence game with the release of a politer (and slightly stupider) machine mind

SERP Abandonment



Neil Postman
TECHNOPOLY



Neil Postman

TECHNOPOLY

*The Surrender of Culture
to Technology*



Daedalus
Information Systems

read by Jeff Rigenbach

unabridged



Key Takeaways

Broaden scope of awareness

Understand the landscape and influences

Embrace new tools and methodologies

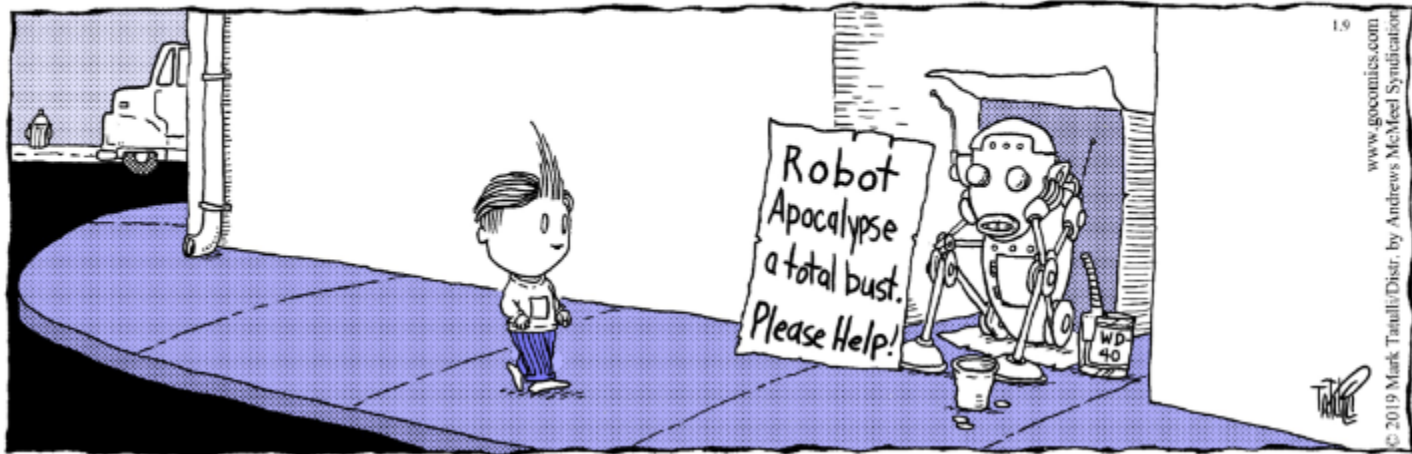


Thank You

Embrace, engage, define, direct

Lio

BY MARK TATULLI



Marianne Sweeny

Principal

Daedalus Information Systems

sweeny48@uw.edu

@msweeny

