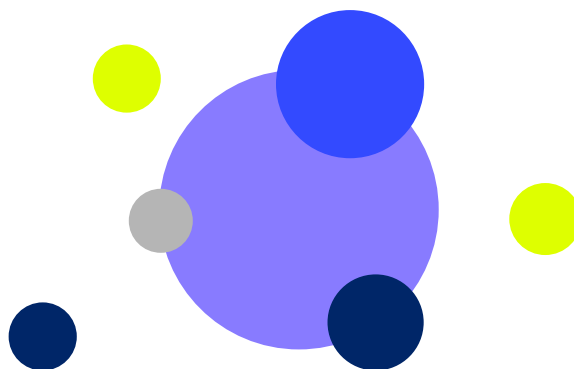**PROPHET**

Innovation + Trust

Building an Ethical Foundation
for Artificial Intelligence

SUSAN ETLINGER
@setlinger

1

Susan Etlinger
Senior Analyst
Altimeter, a Prophet Company

Senior Fellow
Centre for International Governance Innovation

RESEARCH REPORT
Innovation + Trust
The Foundation of Responsible Artificial Intelligence
BY SUSAN ETLINGER
JULY 9, 2018

The AI Maturity
Playbook:
Five Pillars of
Enterprise Success
DECEMBER 19, 2017
BY Susan Etlinger

RESEARCH REPORT
AI in the Enterprise
Real Strategies for
Artificial Intelligence
JUNE 12, 2018
BY SUSAN ETLINGER

RESEARCH REPORT
GDPR Beyond
May 25, 2018
Implications for Strategists
and Marketers
FEBRUARY 26, 2018
BY SUSAN ETLINGER

THE
CONVERSATIONAL
BUSINESS
HOW CHATBOTS WILL
RESHAPE DIGITAL
EXPERIENCES
By Susan Etlinger, Analyst
Altimeter, a Prophet Company
June 27, 2017

2

2

1

Clarifying Terms

1. Artificial Intelligence

Two key attributes distinguish AI from other types of technology: it can act autonomously, and it can learn, sometimes without human intervention.

2. Ethics

Ethics is, generally speaking, a system of moral principles that govern human behavior.

3

According to industry analyst firm Gartner, adoption of artificial intelligence has tripled in the last year alone. An estimated 37 percent of organizations now use AI in some form.

Source: Gartner 2019 CIO survey

4

Here's what's driving that momentum:

1 OPTIMIZATION

Goals
Improve productivity,
reduce costs.

2 IMPROVED
DECISION-MAKING

Improve customer experience,
margin, revenue, market share.

3 BUSINESS MODEL
INNOVATION

Method
Robotic Process Automation,
followed by Natural Language
Processing

AI, Machine Learning

Disrupt own or neighboring industry
to identify new ways of doing busi-
ness, new markets, net revenues,
new business models.

Value
Descriptive

Predictive

AI, Machine Learning
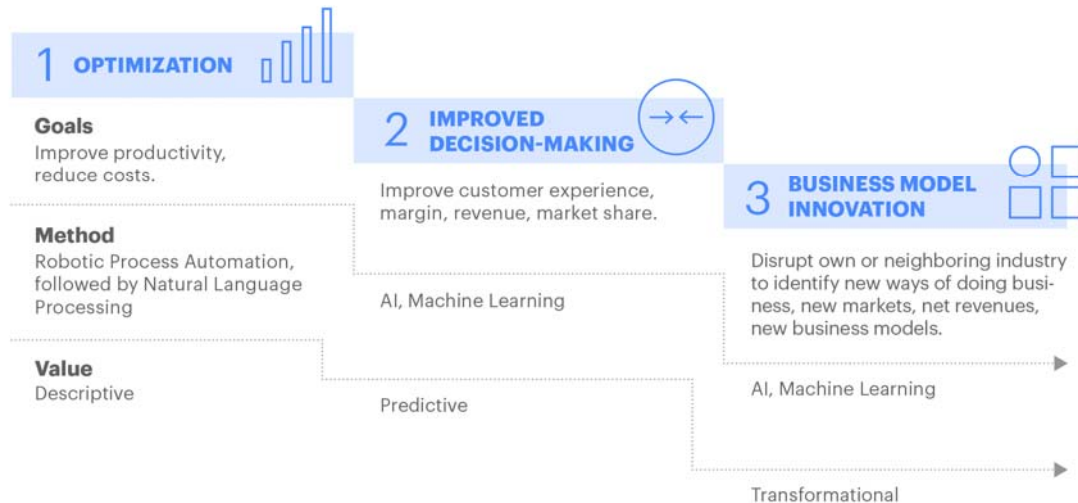
Transformational

5

5

AI is still relatively new. It's complex. It's at times
very overhyped. But it is also legitimately
transformative.

To understand why, it's important to understand
five macro trends shaping this industry.

6

6

Here are the big changes happening around us:

How we *interact*:   from screens to senses
How we *decide*:   from business rules to probabilities
How we *lead*:   from hierarchy-led to data-led
How we *innovate*:   from data as exhaust to data as DNA
How we *behave:*   from the 'wild west' to responsible AI

7

7

How we interact: screens to senses

VISION
Interpreting the objects and attributes of visual objects and images: photos, drawings, video

HEARING
Interpreting sounds or text and translating them into speech, text, or images, or from one language to another

TOUCH
Pinch-and-zoom, gestural interfaces (such as in AR/VR) that translate movement into meaning or commands

SMELL
The ability to translate smells into digital chemical information, and vice-versa

TASTE
The ability to translate taste into digital chemical information, and vice-versa

— Related Science/Technology —

Computer Vision | Speech Recognition, Natural Language Understanding | Gesture-Based Communication, Touch-Based Inputs | Digitizing Olfaction | Digitizing Taste

8

8

How we decide: rules to probabilities

Rules-based (If *x*, then *y*)

If credit card is used outside the United States without a travel plan on file, freeze balance and issue fraud alert.

Probability-based (AI/ML)

Based on historical data, there is a 92% chance that transaction is valid. Authorize based on amount, past transaction history and location history.

9

9

How we innovate: data as exhaust to data as DNA

Research from PwC shows that 45% of total economic gains by 2030 will come from product enhancements, stimulating consumer demand.

This is because AI will drive greater product variety, with increased personalization, attractiveness and affordability over time.
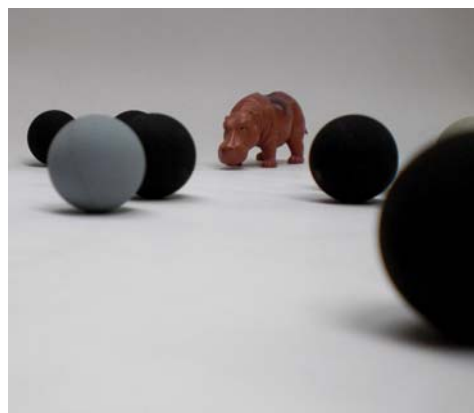
pwc

10

10

How we lead: hierarchy-driven to data-driven

"Leaders need to get comfortable with driving strategy based on what data is telling them. They need to become data-driven versus HiPPO-driven."

Omar Tawakol
Founder and CEO, Voicea

Source: timlewisnm, cc 2.0

11

11

How we behave: the 'Wild West' to responsible AI

Figure 1. Key AI Ethics Issues for Business

1 BIAS
The processes and techniques used to reveal and address societal and statistical bias in data models and algorithms that underlie products and services.

2 EXPLANATION
The ability to provide customers, consumers, citizens, patients, and other constituencies with explanations for decisions made by algorithms.

3 AUTHENTICITY & DISCLOSURE
How organizations use new interaction models, such as chatbots, voice agents, and robots, in a trustworthy manner.

4 GOVERNANCE
Policies and processes that govern the use of AI technology within and outside the organization.

5 IMPACT
The human and business impact of ethical programs, policies, and processes on the organization.

Source: "Ethical AI: A Quick-Start Guide for Executives", Altimeter: March 2019
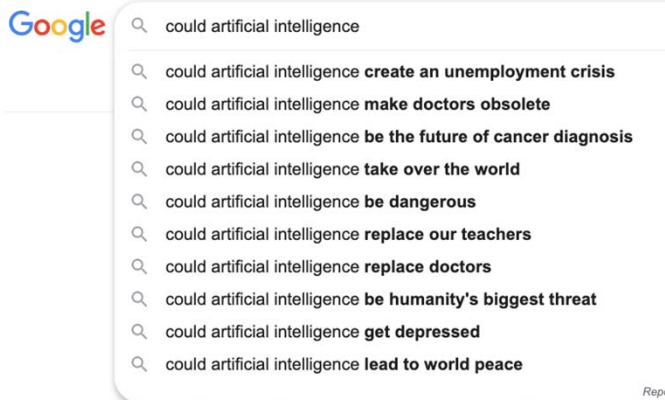
12

12

How we think about the impact of AI

The Silicon Valley

"Nobody of substance at the big companies thinks of algorithms as neutral. Nobody is not aware of the risks."

– Alex Stamos, Former Chief Security Officer, Facebook

The Real World

Google   could artificial intelligence

could artificial intelligence **create an unemployment crisis**
could artificial intelligence **make doctors obsolete**
could artificial intelligence **be the future of cancer diagnosis**
could artificial intelligence **take over the world**
could artificial intelligence **be dangerous**
could artificial intelligence **replace our teachers**
could artificial intelligence **replace doctors**
could artificial intelligence **be humanity's biggest threat**
could artificial intelligence **get depressed**
could artificial intelligence **lead to world peace**

13

Four Ethics Issues Specific to AI

1	**2**	**3**	**4**
BIAS	AUTHENTICITY	EXPLANATION	ETHICAL USE
Documented bias in data/algorithms that results in real harms to health, economic well-being, safety, trust, etc.	New interaction models such as chatbots, voice agents & deepfakes challenge notions of authenticity and transparency.	"The black box": we don't always know why algorithms reach a certain conclusion; especially in proprietary systems.	Few to no governance structures/norms that ensure research, products, services and brand reflect our values.

14

Bias Behaves Like a Toxin

Society
- Education
- Criminal Justice
- Healthcare
- Credit and Insurance
- Government
- Dignity and Self-Worth

Industry
- Recommendation Engines
- Translation
- Audience Segmentation
- Personalization
- Ad Targeting
- Claims Processing
- Voice Agents and Chatbots
- Brand Experience and Trust

Prophet   Proprietary and confidential. Do not distribute.   15

15

Data Encodes Bias

Bias in Language Data

Bias in Voice Data

Extreme he occupations	Extreme she occupations
1. Maestro	1. Homemaker
2. Skipper	2. Nurse
3. Protégé	3. Receptionist
4. Philosopher	4. Librarian
5. Captain	5. Socialite
6. Architect	6. Hairdresser
7. Financier	7. Nanny
8. Warrior	8. Bookkeeper
9. Broadcaster	9. Stylist
10. Magician	10. Housekeeper
11. Fighter Pilot	11. Interior Designer
12. Boss	12. Guidance Counselor

Intelligent Machines

AI Programs Are Learning to Exclude Some African-American Voices

Voice interfaces, chatbots, and other systems are discriminating against certain minority dialects.

by Will Knight   August 16, 2017

Prophet   Proprietary and confidential. Do not distribute.

Sources: "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings": https://arxiv.org/pdf/1607.06520.pdf, MIT Technology Review   16

16

Bias in Computer Vision

"We find that (a) datasets for these tasks contain significant gender bias and (b) models trained on these datasets further amplify existing bias."
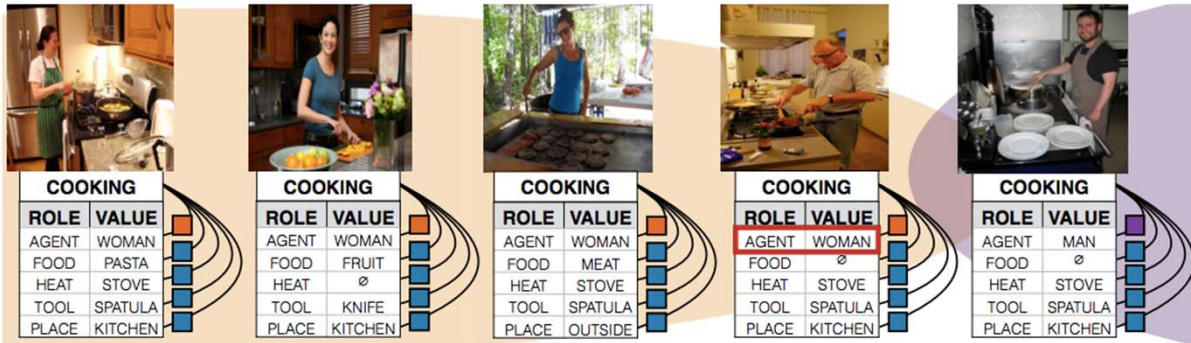
17

17

Bias in Face Recognition

Amazon face recognition falsely matches 28 lawmakers with mugshots, ACLU says

Test of Rekognition software links members of Congress to arrest photos and finds people of color misidentified disproportionately

▲ The civil rights leader John Lewis was among the misidentified members of Congress. Photograph: Jim Lo Scalzo/EPA

Amazon's facial recognition technology falsely identified 28 members of Congress as people who have been arrested for crimes, according to the American Civil Liberties Union (ACLU).

The Gender Shades Project

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE**	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

Sources:: John Lewis Media Center, Joy Buolamwini, Gender Shades Project, MIT

18

18

Use Cases for Facial Recognition

DETECT
Is there a face?

YES   NO

IDENTIFY or VERIFY
Have I seen this face?
Am I looking for this face?

CLASSIFY
What type of face?
(age, gender, emotion, etc.)

(A) Identify: one-to-many
ex. police search

(B) Verify: one-to-one
Is this the same face?
ex. unlock phone, pay with face

Source: Gender Shades Project, MIT   19

19

(Mis)understanding Emotion and Gender

In this image, taken from a research paper on emotion and gender classification, the men (in red) are categorized as "neutral", while the woman (in blue is classified as "angry".

[Source: "Real-time Convolutional Neural Networks for Emotion and Gender Classification", https://arxiv.org/pdf/1710.07557.pdf]
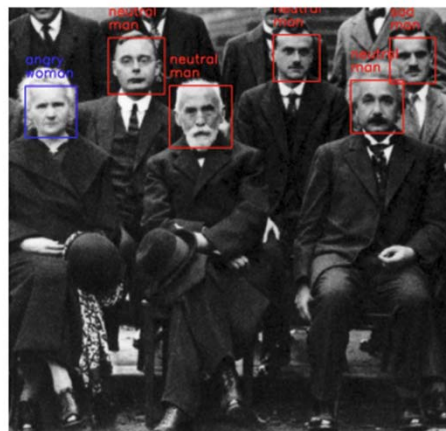
Fig. 6: Results of the provided combined gender and emotion inferences demo. The color blue represents the assigned class *woman* and red the class *man*

20

20

Synthetic Video: "Deepfakes"

21

21

Robots and Transparency

Chatbots   Voice Agents   "Deepfakes"

22

22

From "Black Box" to Explainable AI

AI System
http://explainthatstuff.com

- We are entering a new age of AI applications
- Machine learning is the core technology
- Machine learning models are opaque, non-intuitive, and difficult for people to understand

DoD and non-DoD Applications

Transportation
Security
Medicine
Finance
Legal
Military

User

- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

Figure 1. The Need for Explainable AI

Source: Mr. David Gunning, DARPA

23

23

Governance and Ethical Use

occer US politics Business **Tech** Science Homelessness

New AI can guess whether you're gay or straight from a photograph

An algorithm deduced the sexuality of people on a dating site with up to 91% accuracy, raising tricky ethical questions
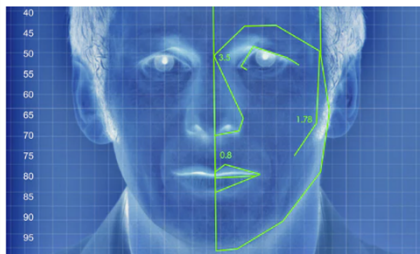
▲ An illustrated depiction of facial analysis technology similar to that used in the experiment. Illustration: Alamy

Artificial intelligence can accurately guess whether people are gay or straight based on photos of their faces, according to new research that suggests machines can have significantly better "gaydar" than humans.

The growing backlash against facial recognition tech

Apple, Amazon, and Microsoft are all mired in controversy over it.

By Sigal Samuel | Apr 27, 2019, 8:00am EDT

A woman holds a tablet featuring biometric 3D facial recognition software. | Getty Images

24

Sources: The Guardian, Vox

24

Good Intentions Aren't Enough

"Even well-intentioned people who are building these models—
left to their own devices or without proper guidance—could
easily end up reproducing or propagating these issues."

– Solon Barocas, Cornell University

Quoted from "Bias Traps in AI", AI Now 2017   25

25

Building a Responsible Culture of Data and AI

Research and Data Science
- GDPR and ethical data use
- Checklists
- Model audits
- Methodologies

Product & Solution Development
- Design thinking
- Ethics engineered into agile process
- Privacy by design
- Designing for empathy
- Social nudges
- Ethics audits

Governance
- Scenario planning
- Explainability
- Feedback loops
- Ethics principles tied to OKRs
- counsel
- Ethics audits
- Ethics education
- Advisory counsels

Culture & Community
- Diverse, empowered employees
- Value-driven KPIs
- Ethics Incentives
- AI ethics research
- Community of practice
- Data science/ Developer practices

26

26

Support, Resources, Education

27

Ethics in Action

"AI can help people become more aware of their own biases. On Airbnb, if a host doesn't accept a guest then he or she is automatically prompted to block out those dates because the system assumes they are no longer available. If the host doesn't, then he or she is asked why."

Queenie Wong, *The Mercury News*

28

Source: Toni Vuohelainen, cc 2.0

29

Four Provocations

1	2	3	4
AI increasingly shapes our (and our customers' and audience's) reality. This will only intensify.	Regulation (such as GDPR and CCPA) will never be able to keep up with technology.	The more *human* AI becomes, the more *humane* it must be.	Ethics and innovation are ***not*** a zero-sum game.

Prophet   Proprietary and confidential. Do not distribute.

30

30

15

Thank you!

Susan Etlinger

susan@altimetergroup.com

@setlinger

LinkedIn: susanetlinger

Altimeter, a Prophet company, provides research and advisory for companies challenged by business disruption, enabling them to pursue new opportunities and business models.

31