Artificial Intelligence: Technology and Industry

Gary Geunbae Lee, POSTECH
Artificial Intelligence: What is AI?
Turing (1950) “Computing machinery and intelligence”:
“Can machines think?” $\rightarrow$ “Can machines behave intelligently?”
Operational test for intelligent behavior: the Imitation Game
What Can AI Do?

Quiz: Which of the following can be done at present?

- ✔ Play a decent game of table tennis?
- ✔ Play a decent game of Jeopardy?
- ✔ Drive safely along a curving mountain road?
- ✔ Drive safely along Telegraph Avenue?
- ✗ Buy a week’s worth of groceries on the web?
- ✗ Buy a week’s worth of groceries at Berkeley Bowl?
- ✔ Discover and prove a new mathematical theorem?
- ✗ Converse successfully with another person for an hour?
- ✔ Perform a surgical operation?
- ✔ Translate spoken Chinese into spoken English in real time?
- ✔ Fold the laundry and put away the dishes?
- ✗ Write an intentionally funny story?
Computer Vision
Autonomous Driving

**Look — no driver**

**Video camera**
Detects traffic lights, oncoming vehicles and other obstacles

**Lidar**
A rotating sensor on the roof scans 200ft in all directions to create a 3D map of its surroundings

**Radar**
Three sensors at the front and one at the back work out the positions of nearby objects

**Position estimator**
A sensor on the left rear wheel measures the car's movements so that its position can be mapped with accuracy
Handwriting Recognition
Sentiment Analysis

“This movie should have NEVER been made. From the poorly done animation, to the beyond bad acting. I am not sure at what point the people behind this movie said "Ok, looks good! Lets do it!" I was in awe of how truly horrid this movie was.”

Positive or Negative?
Machine Translations

The Prime Minister has launched another track - without explaining and many experts at the environmental conference could not do - : the mobilization of some of the financial gains earned on the French nuclear fleet. "Throughout the remaining life of our plants, and while ensuring maximum security, said Jean-Marc Ayrault, our nuclear fleet will be involved without supply disruption."

Le premier ministre a lancé une autre piste - sans l'expliquer et beaucoup des experts présents à la conférence environnementale n'ont pu le faire - : la mobilisation d'une partie des gains financiers perçus sur le parc nucléaire français. "Pendant toute la durée de vie restante de nos centrales, et tout en assurant une sécurité maximale, a déclaré Jean-Marc Ayrault, notre parc nucléaire sera mis à contribution sans rupture d'approvisionnement".
Virtual Assistants
Dialog System

2016 ten breakthrough technology from MIT technology review
Chat bot
Tutor robot
...

Humans versus Machines

1997: Deep Blue (chess)    2011: IBM Watson (Jeopardy!)    2016: AlphaGo
Many AI Applications

- Web search
- Speech recognition
- Handwriting recognition
- Machine translation
- Information extraction
- Document summarization
- Question answering
- Spelling correction
- Image recognition
- 3D scene reconstruction
- Human activity recognition
- Autonomous driving
- Music information retrieval
- Automatic composition
- Social network analysis

- Product recommendation
- Advertisement placement
- Smart-grid energy optimization
- Household robotics
- Robotic surgery
- Robot exploration
- Spam filtering
- Fraud detection
- Fault diagnostics
- AI for video games
- Financial trading
- Dynamic pricing
- Protein folding
- Medical diagnosis
- Medical imaging
Technology

Stephen Hawking warns artificial intelligence could end mankind

By Rory Cellan-Jones
Technology correspondent

© 2 December 2014 Technology

Stephen Hawking: "Humans, who are limited by slow biological evolution, couldn't compete and would be superseded"

The Dawn of the Age of Artificial Intelligence

Reasons to cheer the rise of the machines

Elon Musk: AI Is Going to Happen. Let's Prepare For It

"It's definitely going to happen. So if it's going to happen, what's the best way for it to happen?"

TODAY'S MUST READS

1. Speaking Points That Make You Sound, Like, Totally Unprofessional

2. How Playing the Long Game Made Elizabeth Holmes a Billionaire

3. Holy Knickknacks to Celebrate Pope Francis's Visit

Inside the Mind of Facebook's Sheryl Sandberg

Take a Video Tour of Facebook's Frank Oyer-Designed New York City Office
A (Short) History of AI

§ 1940-1950: Early days
- 1943: McCulloch & Pitts: Boolean circuit model of brain
- 1950: Turing’s “Computing Machinery and Intelligence”

- 1950—70: Excitement: Look, Ma, no hands!
  - 1950s: Early AI programs, including Samuel’s checkers program, Newell & Simon’s Logic Theorist, Gelernter’s Geometry Engine
  - 1956: Dartmouth meeting: “Artificial Intelligence” adopted
  - 1965: Robinson’s complete algorithm for logical reasoning

- 1970—90: Knowledge-based approaches
  - 1969—79: Early development of knowledge-based systems
  - 1980—88: Expert systems industry booms

- 1990—2012: Statistical approaches + subfield expertise
  - Resurgence of probability, focus on uncertainty
  - General increase in technical depth
  - Agents and learning systems... “AI Spring”?

§ 2012—: Excitement: Look, Ma, no hands again?
- Big data, big compute, neural networks
- Some re-unification of sub-fields
- AI used in many industries
Characteristics of AI Tasks

High societal impact (affect billions of people)

Diverse (language, games, robotics)

Complex (really hard)
Two sources of complexity

**Computational Complexity**
- Most AI problems are NP-hard
- Go – 361,200 trajectories that a player would have to consider to play optimally.

**Information Complexity**
- Translate a sentence
- Classify a bird from image
Resources

Computation (time/memory)  Information (data)
Big Data + Hardware + Machine Learning Algorithm

AlphaGo (2016)
• 30 million training data
• Tensor Processing Unit
• Deep & Reinforcement Learning

NVIDIA Self-Driving (2016)
• Vision data by driving tens of thousands miles
• NVIDIA GPU, Deep Learning

Google Translation
• Billions of translation data
• IBM’s linguistic approach fails
Real-world task

Modeling

Formal task (model)

Algorithms

Program
Algorithms (example)

• Formal task:
  • Input: list $L = \{x_1, \ldots, x_n\}$ and a function $f : X \rightarrow \mathbb{R}$
  • Output: $k$ highest-scoring elements

• Example ($k = 2$):
  • $L : A B C D$
  • $f : 3 2 7 1$

• Two algorithms:
  1. Scan through to find the largest, scan through again to find the second largest, etc.
  2. Sort $L$ based on $f$, return first $k$ elements
Modeling (example)

• Real-world task:
  • Input: 20 billion web pages, a keyword query
  • Output: 10 most relevant web pages

• Modeling:
  • $L = \text{list of web pages}$
  • $f(x) = 10 \times \text{QueryMatch}(x) + 3 \times \text{PageRank}(x)$

• Formal task:
  • Input: list $L = \{ x_1, \ldots, x_n \}$ and a function $f : X \rightarrow \mathbb{R}$
  • Output: $k$ highest-scoring elements
Machine learning

Reflex

States

Variables

Logic

Search problems
Markov decision processes
Adversarial games

Constraint satisfaction problems
Bayesian networks

"Low-level intelligence"

"High-level intelligence"
Artificial Intelligence: Industrial Application
AI vs. ML vs. DL

**Artificial Intelligence**
- Perception (Vision, Speech, NLP)
- Reasoning/Planning/Data Analytics
- Knowledge Representation
- Learning
- Manipulation/Navigation

**Machine Learning**
- Supervised Learning with Teachers
- Semi-supervised Learning with Teacher
- Unsupervised Learning without Teacher
- Reinforcement Learning with Rewards

**Deep Learning**
- RBM (Restricted Boltzmann Machine)
- DBN (Deep Belief Network)
- CNN (Convolutional Neural Network)
- Deep Reinforcement Learning

**Input**
- Text
- Sound
- Image
- Video

**Output**
- Generated Information
- / Autonomous Control
Venture Investing in Artificial Intelligence

- Total venture funding in our Artificial Intelligence sector

  - Tracking over 852 Future of Tech Venture companies in 13 categories across 62 countries, with a total of $2.71 Billion in funding
### Aspects of AI-driven products and services

<table>
<thead>
<tr>
<th>Ambience</th>
<th>Automaticity</th>
<th>Personalization</th>
<th>Natural Interaction</th>
<th>Self-Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI provides services as ubiquitous and ambient to the user</td>
<td>Minimizing human intervention for automating tasks</td>
<td>AI understands user to provide personalized services</td>
<td>Human-level naturalness as Multi-modal I/F</td>
<td>AI can improve the ability through self-learning from data</td>
</tr>
<tr>
<td><strong>Forecast user problems and solve them with AI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DeepClint</strong></td>
<td><strong>Automatic Object Tagging</strong></td>
<td><strong>Recommending favorite restaurants, places based on user preference/ intention</strong></td>
<td><strong>Converts text to natural sounding voice</strong></td>
<td><strong>Fraud detection and prevention by large-scale ML system</strong></td>
</tr>
<tr>
<td>Criminal Detection, Investigation and Prevention by AI security</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>InS1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Images of AI-driven products and services](image-url)
New Biz opportunity driven by AI technology

- Key use cases and biz opportunities to be enabled by AI technology

<table>
<thead>
<tr>
<th>Personal Affective Agent</th>
<th>Shopping Assistant</th>
<th>Device Enrichment</th>
<th>24/7 After-sales Care</th>
<th>Intelligent IoT Hub</th>
<th>Personal Health Advisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robot/Assistant</td>
<td>Commerce</td>
<td>Smart Device</td>
<td>Customer Service</td>
<td>Smart Home</td>
<td>Healthcare</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social Robot</th>
<th>Driving Assistant</th>
<th>Augmented HMD</th>
<th>Errand Robot</th>
<th>Personalized Car Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robot/Assistant</td>
<td>Automotive</td>
<td>AR</td>
<td>Robot/Assistant</td>
<td>Automotive</td>
</tr>
</tbody>
</table>

Key use cases
Biz opportunities
AI for Personal Affective Agent

- From Engadget report, “Personal assistants are ushering in the age of AI at home”

Personal assistant understands user’s need and situation, then replacing mental labors for the user autonomously

- Phone calls, SMS, emails
- Schedule, contacts, social information
- Voice commands, Messenger / chatting
- Voice and Text Input

Extracting device contextual info

Understanding the place and user mood

Analyzing user behavior and social information

- Detect patterns
- Analyze schedule
- Infer relationships

Predicting Actions

User Intent Model

Understanding Context

Predicting the next action and user’s goal

- Usage Logs
- Status Monitoring
- Location, weather, To do list

Automatic schedule/contacts management

Seamless agent for any device and environment

Automatic app operation based on the situation and user activity
AI for Personal Health Advisor

- Healthcare assistant manages the user health status by providing personalized medical information and advise by analyzing user behavior

- From Medical Futurist report, “Artificial Intelligence Will Redesign Healthcare”
AI for Shopping Assistant

- Commercial assistant provides automatic search and discovery over products, and recommends the best matches timely for users.

- From Information Age report, “3 ways artificial intelligence is transforming e-commerce.”
AI for 24/7 After-sales Care

- From Wikipedia, “Customer support Automation”

Customer service assistant provides preventive diagnosis, failure prediction and automate customer support via knowledge base and expert system.

Preventive Customer Care

- Home-1
- Home-2
- Home-3
- Home-4

Refrigerator might be out of order in 30 days.

Do you want to reserve customer service?

24/7 A/S Care

- Diagnosis logs
- Root-cause modeling
- AI based CS agent

Customer Service Automation

- Voice call agent
- SNS agent

Human backend
Driving assistant provides safety-enhanced sensing and control via semi-automatic driving mode

- From TechCrunch report, “Automatic, a smart driving assistant, could be the next big thing”
AI for Intelligent IoT Hub

- Contextual services by understanding users and environments via IoT devices/sensors network

- From WIRED report, “IoT won’t work without Artificial Intelligence”
AI for Errand Robot

- Errand robot provides real-world/human-scale physical actions to replace various manual labors for users

- From NBC NEWS, “A robot to run your errands”
AI for Social Robot

- Social robot understands user’s emotion and mood to provide social relationship roles and home/family care

- Emotional interaction
  - Emotional communication via conversation, touch and feel

- Toddler/Senior Care
  - Entertainment and education for toddlers and seniors

- Social entertainment robot
  - Affective computing, action/status/mood recognition
  - Interacts with user’s emotion and feelings

- Butler robot
  - AI Home manager monitors home events and housekeeping with map

- Family/Home Agent
  - User preference, personalized content and information for family members

- From Wikipedia, “Social robot”
Artificial Intelligence
Technology and Trend
Artificial Neural Networks (ANN)

Σ

Activation Function

Dendrites

Terminal Branches of Axon

\[ \sum x_1 x_2 w_1 w_2 w_3 \ldots w_n xn \]
Layered Networks

\[
\sum_{i,j} = f(w_i x)
\]
\[
y = f(w_i x + w_j x + w_k x + \cdots + w_m x_m)
\]

Output: \( y = f(\sum_{j} w_i^j x_j) \)
Deep learning Innovation

- Combining Feature Learning and Classification as Unified Framework (※ Learning what to learn, how to learn)
Vanilla recurrent neural networks (RNNs)

- RNNs have connections from the outputs of previous time steps to inputs of next time steps.

- For sequential data, a RNN usually computes hidden state $h_t$ from the previous hidden state $h_{t-1}$ and the input $x_t$:
  - $h_t = \sigma(W_h h_{t-1} + W_x x_t + b)$
Long short-term memory networks (LSTMs)

• LSTMs explicitly keep and update cell memory $c^{(t)}$ by
  • Removing the previous cell content $c^{(t-1)}$ by multiplying it with $f^{(t)}$
  • Adding the new cell content $\tilde{c}^{(t)}$ multiplied by $i^{(t)}$
• LSTMs produce output $h^{(t)} = o^{(t)} \circ \tanh c^{(t)}$

\[
\begin{align*}
  f^{(t)} &= \sigma \left( W_f h^{(t-1)} + U_f x^{(t)} + b_f \right) \\
  i^{(t)} &= \sigma \left( W_i h^{(t-1)} + U_i x^{(t)} + b_i \right) \\
  o^{(t)} &= \sigma \left( W_o h^{(t-1)} + U_o x^{(t)} + b_o \right) \\
  \tilde{c}^{(t)} &= \tanh \left( W_c h^{(t-1)} + U_c x^{(t)} + b_c \right) \\
  c^{(t)} &= f^{(t)} \circ c^{(t-1)} + i^{(t)} \circ \tilde{c}^{(t)} \\
  h^{(t)} &= o^{(t)} \circ \tanh c^{(t)}
\end{align*}
\]
Bidirectional Multi-Layer RNNs
• CNN architecture for sentence classification

Transformer

- Parallel self-attention
  - Looks at self, and determines where to focus

\[
\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V
\]

Q, K, V – vectors for every word, output attention summed up; head means different Q, K, V vector with different weights.

Vaswani, Ashish, et al. "Attention is all you need." NIPS
• Training 1. Masked words prediction
  • 15% of words are [MASK]ed

*GELU: Gaussian error linear unit
BERT: Bidirectional Encoder Representations from Transformers

- BERT as universal pre-trained model for NLP
  - BERT requires minimal additional layers and fine-tuning

*GLUE benchmark task

In pre-training, optimize $L_1(u)$

$u$: Unlabeled dataset

$\Theta$: Model parameters

In fine-tuning, optimize $L_3(c)$

$c$: Labeled dataset

$\lambda$: Hyper-parameter weight
Distributed/ Parallel computing for Deep Learning

• History of parallel/ distributed systems for Deep Learning computing

Google taps 16k computers to look for cats – for Science!

Univ. of Toronto uses 2 GPUs for 1.2M training Images for 1000 classes Image classification (※ ImageNet Large Scale Visual Recognition Challenge)

Stanford uses 12 GPUs for Large-scale Video Classification With Convolutional Neural Networks (※ 10M Youtube video)

Google uses 16K CPU cores for Training 22-layers Deep neural network (※ GoogLeNet, 2014)

Baidu’s Artificial Intelligence Supercomputer Beats Google at Image Recognition
Deep Learning for Computer Vision

- Examples of vision problem with deep learning

**Prediction (Generation)**
- Image Generation
- Frame Prediction
- Sound generation from Video

**Detection, Segmentation**
- Object Detection
- Object Tracking
- Video Search

**Classification**
- Image Classification
- Activity Recognition
- Video Search

**Image**

- The man at bat readies to swing at the pitch while the umpire looks on

**Video**

- Answer phone
- Drive car

**Multi-Modal**

- Image Captioning
Caption generation from image/video

- Text description for image/video contents
  - Input: image/video, output: text description
  - Need to learn visual and language feature

Image captioning (from Vinyals et al.)

A group of people shopping at an outdoor market

Video description (from Donahue et al.)

A man is juicing the orange
End to end learning for steering control

- Steering control for autonomous driving
  - End to End learning: from camera image to steering command (steering wheel angle)
  - Trained to minimize error between the steering command output by network and human.

**Training architecture**

**Running architecture**
Driving situation understanding

- Understand road environment
  - Detect car, pedestrian, road lane, traffic sign
  - Segment safe driving area
  - On device real-time processing is needed

Object detection with multi-layer input (from Shrivastava et al.)
Action Detection from Video

- Detect action [action class, frame, object area] from videos
  - Spatial CNN for proposal detection
  - Motion CNN with optical flow for movement tracking
  - Both spatial and motion information is used to classify action

Action Detection Architecture (from Gkioxari et al.)
Image(Sample) generation with GAN

- Realistic generation of samples by generative model
  - Both generative model (sample generation) and discriminative model (classifier) is used.
  - It can be used to predict missing input for semi supervised learning

GAN Concepts

GAN examples
Visual Product Search: Product recognition

- Challenges in product recognition

**Large scale**
Recognizing >1M objects

**One-shot**
Each product has one or few examples

**Unseen objects**
New product every day

[ example of product DB ]

Label information is unavailable during training phase
Visual Product Search : Deep Feature (metric learning)

- Learning feature embedding: learn such that similar objects are located on nearby points
- High-dimensional feature indexing: fast search in large scale data

① Training Data
② Deep Network Training
③ Product DB
④ Deep Feature Extraction
⑤ Image Query
⑥ Deep Feature Search
⑦ Search Result
Personal Assistant Big Picture

- Building human-like personal companion who knows and understands you better than yourself
Deep Learning in Personal Assistant

• Definition
  – Based on the user’s utterance, doing the proper action or finding the right answers

• Architecture

Area which could be applied Machine Learning/ Deep Learning
  – Each Component (e.g. ASR, NLU, etc.)
  – End-to-end Learning based on one neural network architecture
    • NLU ~ NLG
End-to-end ASR

Frontend (Preprocessing) STFT, MEL

Audio Inputs

Transformer Encoder

Speech Representation

Transformer Decoder (Attention)

CTC

Beam Search

Text Outputs

*Connectionist Temporal Classification (CTC)
Tacotron2: Seq2seq with attention RNN + modified WaveNet

*waveNet: invert the mel spectrogram feature representation into time-domain
waveform samples (10 component mixture of logistic distributions (MoL) to generate 16-bit
samples at 24 kHz)

*location-sensitive attention: mitigating potential failure modes where some subsequences are repeated or ignored by the decoder

*Auto-regressive decoder to generate mel spectrogram frame by frame; Previous frame input to pre-net & post-net (for residual)

*stop token to stop generation
Natural Language Understanding with Deep Learning

- NLU is to find the user’s intent and slots from the user’s utterance text (e.g. “find the restaurant in Seoul” -> intent:find.restaurant, slot:”Seoul”)
  - Finding Intent and Slots based on Joint Model
  - Utilizing previous user’s utterances based on Memory Network
• NLU module is shared for DST, POL, and NLG
• Darker blocks mean previous turn
• DB result contains the number of matched entries for each domain

Cross entropy loss

Binary cross entropy loss
User: How are things?
Meena: They’re good. How about you?
User: I’m trying to decide what to watch
Meena: What are some of your favorite shows?
User: I like the Good Place
Meena: I haven’t watched that. My favorite shows are Supernatural, and Star trek: TNG.
User: What is TNG?
Meena: The Next Generation

*Automatically learned architecture w/ NAS (neural architecture search)

Text Input Prediction for Keyboard

- Text Prediction for Keyboard Users
  - Current Word Completion
  - Next Word Prediction
  - Typo Auto Correction
  - Continuous Input Detection (Swype)

- Deep Learning based Language Modeling
  - LSTM based model to understand contexts
  - Embedded Deep Learning Platform
Summary

• Applications of AI: high-impact, diverse

• Challenges: computational/information complexity

• Paradigm: modeling + algorithms

• Models: learning + [reflex, states, variables, logic]
Q&A