**Experiment 10**

Strong artificial intelligence (AI), also known as artificial general intelligence (AGI) or general AI, is a theoretical form of AI used to describe a certain mindset of AI development. If researchers can develop Strong AI, the machine would require an intelligence equal to humans; it would have a self-aware consciousness that can solve problems, learn, and plan for the future.

Strong AI aims to create intelligent machines that are indistinguishable from the human mind. But just like a child, the AI machine would have to learn through input and experiences, constantly progressing and advancing its abilities over time.

While AI researchers in both academia and private sectors are invested in the creation of artificial general intelligence (AGI), it only exists today as a theoretical concept versus a tangible reality.

While some individuals, like Marvin Minsky, have been quoted as being overly optimistic about what we could accomplish in a few decades in the field of AI; others would say that Strong AI systems cannot even be developed. Until the measures of success, such as intelligence and understanding, are explicitly defined, they are correct in this belief. For now, many use the Turing test to evaluate the intelligence of an AI system.

## **Tests of Strong AI**

### *Turing Test*

Alan Turing developed the Turing Test in 1950 and discussed it in his paper, “Computing Machinery and Intelligence”. Originally known as the Imitation Game, the test evaluates if a machine’s behavior can be distinguished from a human's. In this test, there is a person known as the “interrogator” who seeks to identify a difference between computer-generated output and human-generated ones through a series of questions. If the interrogator cannot reliably discern the machines from human subjects, the machine passes the test. However, if the evaluator can identify the human responses correctly, then this eliminates the machine from being categorized as intelligent.

While there are no set evaluation guidelines for the Turing Test, Turing did specify that a human evaluator will only have a 70% chance of correctly predicting a human vs computer-generated conversation after 5 minutes. The Turing Test introduced general acceptance of the idea of machine intelligence.

However, the original Turing Test only tests for one skill set — text output or chess as examples. Strong AI needs to perform a variety of tasks equally well, leading to the development of the Extended Turing Test. This test evaluates the textual, visual, and auditory performance of the AI and compares it to human-generated output. This version is used in the famous Loebner Prize competition, where a human judge guesses whether the output was created by a human or a computer.

### *Chinese Room Argument (CRA)*

The Chinese Room Argument was created by John Searle in 1980. In his paper, he discusses the definition of understanding and thinking, asserting that computers would never be able to do this. This excerpt from his paper, from Stanford’s website, summarizes his argument well,

“Computation is defined purely formally or syntactically, whereas minds have actual mental or semantic contents, and we cannot get from syntactical to the semantic just by having the syntactical operations and nothing else…A system, me, for example, would not acquire an understanding of Chinese just by going through the steps of a computer program that simulated the behavior of a Chinese speaker (p.17).”

The Chinese Room Argument proposes the following scenario:

Imagine a person, who does not speak Chinese, sitting in a closed room. In the room, there is a book with Chinese language rules, phrases, and instructions. Another person, who is fluent in Chinese, passes notes written in Chinese into the room. With the help of the language phrasebook, the person inside the room can select the appropriate response and pass it back to the Chinese speaker.

While the person inside the room was able to provide the correct response using a language phrasebook, he or she still does not speak or understand Chinese; it was just a simulation of understanding through matching questions or statements with appropriate responses. Searle argues that Strong AI would require an actual mind to have consciousness or understanding. The Chinese Room Argument illustrates the flaws in the Turing Test, demonstrating differences in definitions of artificial intelligence.

## **Strong AI vs. Weak AI**

Weak AI, also known as narrow AI, focuses on performing a specific task, such as answering questions based on user input or playing chess. It can perform one type of task, but not both, whereas Strong AI can perform a variety of functions, eventually teaching itself to solve new problems. Weak AI relies on human interference to define the parameters of its learning algorithms and to provide the relevant training data to ensure accuracy. While human input accelerates the growth phase of Strong AI, it is not required, and over time, it develops a human-like consciousness instead of simulating it, like Weak AI.

**Strong AI trends**

While there are no clear examples of strong artificial intelligence, the field of AI is rapidly innovating. Another AI theory has emerged, known as artificial superintelligence (ASI), super intelligence, or Super AI. This type of AI surpasses strong AI in human intelligence and ability. However, Super AI is still purely speculative as we have yet to achieve examples of Strong AI.

With that said, there are fields where AI is playing a more important role, such as:

* Cybersecurity: Artificial intelligence will take over more roles in organizations’ cybersecurity measures, including breach detection, monitoring, threat intelligence, incident response, and risk analysis.
* Entertainment and content creation: Computer science programs are already getting better and better at producing content, whether it is copywriting, poetry, video games, or even movies. OpenAI’s GBT-3 text generation AI app is already creating content that is almost impossible to distinguish from a copy that was written by humans.
* Behavioral recognition and prediction: Prediction algorithms will make AI stronger, ranging from applications in weather and stock market predictions to, even more interesting, predictions of human behavior. This also raises questions about implicit biases and ethical AI. Some AI researchers in the AI community are pushing for a set of anti-discriminatory rules, which is often associated with the hashtag #responsibleAI.

## **Strong AI terms and definitions**

The terms artificial intelligence, machine learning, and deep learning are often used in the wrong context. These terms are frequently used in describing Strong AI, so it’s worth defining each term briefly:

Artificial intelligence defined by John McCarthy is "the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable."

Machine learning is a sub-field of artificial intelligence. Classical (non-deep) machine learning models require more human intervention to segment data into categories (i.e. through feature learning).

Deep learning is also a sub-field of machine learning, which attempts to imitate the interconnectedness of the human brain using neural networks. Its artificial neural networks are made up of layers of models, which identify patterns within a given dataset. They leverage a high volume of training data to learn accurately, which subsequently demands more powerful hardware, such as GPUs or TPUs. Deep learning algorithms are most strongly associated with human-level AI.

## **Deep learning applications**

Deep learning can handle complex problems well, and as a result, it is utilized in many innovative and emerging technologies today. Deep learning algorithms have been applied in a variety of fields. Here are some examples:

* Self-driving cars: Google and Elon Musk have shown us that self-driving cars are possible. However, self-driving cars require more training data and testing due to the various activities that it needs to account for, such as giving right of way or identifying debris on the road. As the technology matures, it’ll then need to get over the human hurdle of adoption as polls indicate that many drivers are not willing to use one.
* Speech recognition: Speech recognition, like AI chatbots and virtual agents, is a big part of natural language processing. Audio input is much harder to process for an AI, as so many factors, such as background noise, dialects, speech impediments, and other influences can make it much harder for the AI to convert the input into something the computer can works with.
* Pattern recognition: The use of deep neural networks improves pattern recognition in various applications. By discovering patterns of useful data points, the AI can filter out irrelevant information, draw useful correlations and improve the efficiency of big data computation that may typically be overlooked by human beings.
* Computer programming: Weak AI has seen some success in producing meaningful text, leading to advances in coding. Just recently, OpenAI released GPT-3, open-source software that can write code and simple computer programs with very limited instructions, bringing automation to program development.
* Image recognition: Categorizing images can be very time-consuming when done manually. However, special adaptions of deep neural networks, such as DenseNet, which connects each layer to every other layer in the neural network, have made image recognition much more accurate.
* Contextual recommendations: Deep learning apps can consider much more context when making recommendations, including language understanding patterns and behavioral predictions.
* Fact-checking: The University of Waterloo recently released a tool that can detect fake news by verifying the information in articles by comparing it with other news sources.

**CASE STUDY**

**DeepFace**

*“Closing the Gap to Human-Level Performance*

*in Facial Verification”*

**Abstract**

Face recognition in unconstrained images is at the forefront of the algorithmic perception revolution. The social and cultural implications of face recognition technologies are far-reaching, yet the current performance gap in this domain between machines and the human visual system serves as a buffer from having to deal with these implications. The Case study presents a system (DeepFace) that has closed the majority of the remaining gap in the most popular benchmark in unconstrained face recognition and is now at the brink of human-level accuracy. It is trained on a large dataset of faces acquired from a population vastly different than the one used to construct the evaluation benchmarks, and it can outperform existing systems with only very minimal adaptation.

Moreover, the system produces an extremely compact face representation, in sheer contrast to the shift toward tens of thousands of appearance features in other recent systems. In modern face recognition, the conventional pipeline consists of four stages:

**detect ⇒ align ⇒ represent ⇒ classify.** They revisited both the alignment step and the representation step by employing explicit 3D face modeling to apply a piecewise affine transformation and derive a face representation from a nine-layer deep neural network.

This deep network involves more than 120 million parameters using several locally connected layers without weight sharing, rather than the standard convolutional layers. Thus, they have trained it on the largest facial dataset to date, an identity-labeled dataset of four million facial images belonging to more than 4,000 identities. The learned representations coupling the accurate model-based alignment with the large facial database generalize remarkably well to faces in unconstrained environments, even with a simple classifier. Their method reaches an accuracy of 97.35% on the Labeled Faces in the Wild (LFW) dataset, reducing the error of the current state of the art by more than 27%, closely approaching human-level performance.

**Environment Analysis**

As every conventional AI has an environment, the AI in our case study has a very unique AI which stands out and may help you in understanding the working of the AI better. The nature of the Environment is as follows:

* Semi-Observable Environment

The AI only takes inputs provided by the user, it does not have any sensors of its own to take real-time data. It has a pre-trained model which can be tweaked.

* Stochastic

The DeepFace Ai has a Stochastic Environment that provides the randomness of the deep face generation. The stochastic environment is random which is not unique and cannot be completely determined by the agent.

* Collaborative

Its environment is also collaborative as it still learns from the data it constantly fetches from other platforms and aims at being the best AI. An agent is said to be in a collaborative environment when multiple agents cooperate to produce the desired output.

* Single-agent

The whole AI works on its own for the generation of the result hence, its environment is referred to Single agent. An environment consisting of only one agent is said to be a single-agent environment.

* Dynamic

The environment is also Dynamic, it keeps changing and updating and aims toward accuracy and precision. Also, the result generated is constantly in motion and that is the most unique part of it. An environment that keeps constantly changing itself when the agent is up with some action is said to be dynamic.

* Continuous

The result generated is constantly in motion and that is the most unique part of it.

Hence, the environment has to constantly change and perform actions for the same.0

The environment in which the actions are performed cannot be numbered i.e. is not discrete and is said to be continuous.

* Sequential

The decisions/actions taken by the AI should be smooth enough to showcase a good motion, hence it needs to be a sequential environment system.

In a Sequential environment, the previous decisions can affect all future decisions.

The next action of the agent depends on what action he has taken previously and what action he is supposed to take in the future.

* Unknown

The Environment is also referred to as an Unknown Environment. In the case of an unknown environment, for an agent to make a decision, it has to gain knowledge about how the environment works.

**Agent Analysis**

The Agent in our AI is a hybrid type consisting of -

Learning agents & Utility-based agents.

**Learning Agent**

A learning agent in AI is the type of agent that can learn from its past experiences, or has learning capabilities.

It starts to act with basic knowledge and then can act and adapt automatically through learning.

A learning agent has mainly four conceptual components, which are:

* 1. Learning element:   
     It is responsible for making improvements by learning from the environment.
  2. Critic:   
     The learning element takes feedback from the critic which describes how well the agent is doing concerning a fixed performance standard.
  3. Performance element:   
     It is responsible for selecting external action
  4. Problem Generator:   
     This component is responsible for suggesting actions that will lead to new and informative experiences.

Hence, learning agents can learn, analyze performance, and look for new ways to improve performance.

**Utility-based agents**

These agents are similar to the goal-based agent but provide an extra component of utility measurement which makes them different by providing a measure of success at a given state.

Utility-based agent act based not only on goals but also on the best way to achieve the goal. The Utility-based agent is useful when there are multiple possible alternatives, and an agent has to choose to perform the best action.

The utility function maps each state to a real number to check how efficiently each action achieves its goals.

**Methodology**

In recent years, computer vision literature has attracted many research efforts in descriptor engineering. Such descriptors when applied to face recognition, mostly use the same operator for all locations in the facial image. Recently, as more data has become available, learning-based methods have started to outperform engineered features, because they can discover and optimize features for the specific task at hand. Here, we learn a generic representation of facial images through a large deep network.

**DNN Architecture and Training:** We train our DNN on a multi-class face recognition task, namely to classify the identity of a face image.

**Normalization:** As a final stage we normalize the features to be between zero and one to reduce the sensitivity to illumination changes.

**The SFC dataset:** It includes 4.4 million labeled faces from 4,030 people each with 800 to 1200 faces, where the most recent 5% of face images of each identity are left out for testing. This is done according to the images’ time-stamp to simulate continuous identification through aging. A large number of images per person provides a unique opportunity for learning the invariance needed for the core problem of face recognition.

**The LFW dataset:** It consists of 13,323 web photos of 5,749 celebrities which are divided into 6,000 face pairs in 10 splits.

**The YTF dataset:** It collects 3,425 YouTube videos of 1,595 subjects (a subset of the celebrities in the LFW). These videos are divided into 5,000 video pairs and 10 splits and used to evaluate the video-level face verification.

**Training on the SFC:** We first train the deep neural network on the SFC as a multi-class classification problem using a GPU-based engine, implementing the standard back-propagation on feedforward nets by stochastic gradient descent (SGD) with momentum (set to 0.9).

**Results on the LFW dataset:** The vision community has made significant progress on face verification in unconstrained environments in recent years. The mean recognition accuracy on LFW [18] marches steadily towards the human performance of over 97.5%

**Ensembles of DNNs:** Next, they have combined multiple networks trained by feeding different types of inputs to the DNN:

1) The network DeepFace-single described above is based on 3D-aligned RGB inputs

2) The gray-level image plus image gradient magnitude and orientation

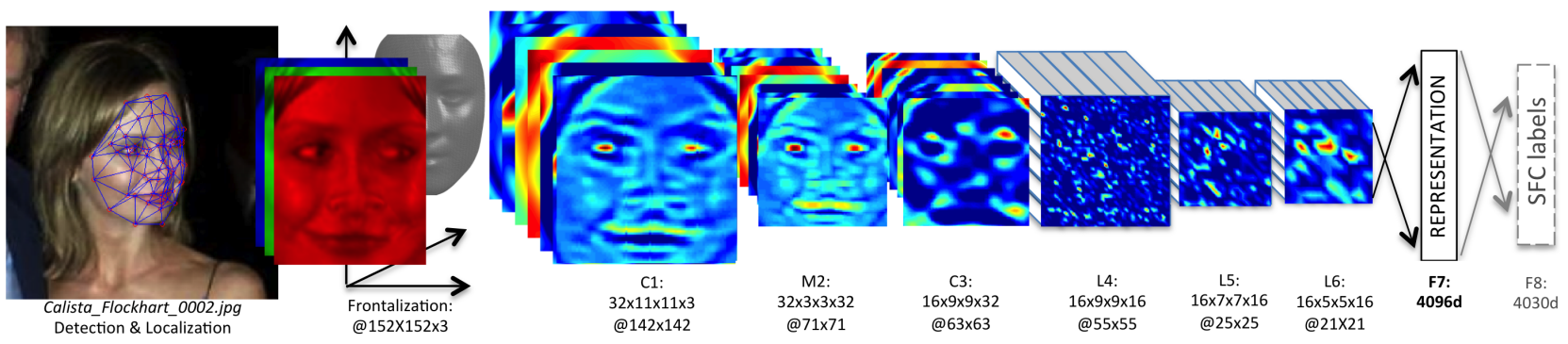
3) the 2D-aligned RGB images.

**Results on the YTF dataset:**

We further validate DeepFace on the recent video-level face verification dataset.

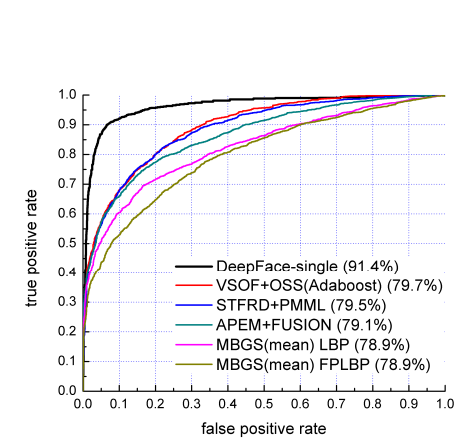
We report an accuracy of 91.4% which reduces the error of the previous best methods by more than 50%.

**Diagrammatic representation**



**Outline of the DeepFace architecture.**

A front-end of single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate feature maps produced at each layer.



**Figure on left:**

The ROC curves on the

YTF dataset.

Best viewed in color.

It collects 3,425 YouTube videos of 1,595 subjects (a subset of the celebrities in the LFW). These videos are divided into 5,000 video pairs and 10 splits and used to evaluate the video-level face verification.

**Computational efficiency**

They have efficiently implemented a CPU-based feedforward operator, which exploits both the CPU’s Single Instruction Multiple Data (SIMD) instructions and its cache by leveraging the locality of floating-point computations across the kernels and the image.

Using a single-core Intel 2.2GHz CPU, the operator takes 0.18 seconds to extract features from the raw input pixels. Efficient warping techniques were implemented for alignment; alignment alone takes about 0.05 seconds.

Overall, the DeepFace runs at 0.33 seconds per image, accounting for image decoding, face detection and alignment, the feedforward network, and the final classification output.

**Summary**

An ideal face classifier would recognize faces with accuracy that is only matched by humans. The underlying face descriptor would need to be invariant to pose, illumination, expression, and image quality.

It should also be general, in the sense that it could be applied to various populations with little modifications if any at all. In addition, short descriptors are preferable, and if possible, sparse features. Certainly, rapid computation time is also a concern.

They believe that this work, which departs from the recent trend of using more features and employing a more powerful metric learning technique, has addressed this challenge, closing the vast majority of this performance gap. Their work demonstrates that coupling a 3D model-based alignment with large capacity feedforward models can effectively learn from many examples to overcome the drawbacks and limitations of previous methods.

The ability to present a marked improvement in face recognition attests to the potential of such coupling to become significant in other vision domains as well.

**Applications**

Face recognition is the state of the art. Face recognition error rates have decreased over the last twenty years by three orders of magnitude when recognizing frontal faces in still images taken in consistently controlled (constrained) environments. Many vendors deploy sophisticated systems for the application of border control and smart biometric identification.   
 However, these systems have shown to be sensitive to various factors, such as lighting, expression, occlusion, and aging, that substantially deteriorate their performance in recognizing people in such unconstrained settings. Most current face verification methods use hand-crafted features.

Moreover, these features are often combined to improve performance, even in the earliest LFW contributions. The systems that currently lead the performance charts employ tens of thousands of image descriptors. In contrast, our method is applied directly to RGB pixel values, producing a very compact yet sparse descriptor.

***\*\*\*\*\****