TECHNOLOGY BRIEFING ON AI INNOVATION AND FOUNDATIONS + USE CASES

Artificial Intelligence (AI) is a branch of computer science dedicated to creating systems capable of performing tasks that, until recently, required human intelligence. These tasks include problem-solving, understanding language, and visual perception.

A common misconception is that the term AI is confined to high-functioning systems like ChatGPT. In reality, it can be as simple as a pre-programmed chatbot, as complex as Google or TikTok's content recommendation system, or as consequential as disease or recidivism risk prediction algorithms. While the connotation of "AI" varies across academia and industry, it has more recently been colloquialized to refer to systems that leverage machine learning or other advanced probabilistic and statistical techniques–which will be discussed throughout this briefing–to provide automotive, generative, or predictive outputs.

Data Collection and Use in AI

Throughout this briefing, we'll use the metaphor of a sculpture to describe AI. With this in mind, data is the clay. At its core, data is the representation of facts or information. While traditionally we might think of data as numbers or text, the digital age has expanded this to include multimedia forms like images, audio, and video. This vast and varied nature of data feeds into the versatility and potential of AI systems.

Types of Data

- **Structured Data**: This is data that adheres to a specific model or structure. Imagine a library catalog where every book is listed with its title, author, publisher, and publication date in a specific order.
  - *Example*: Relational databases, like those managed using SQL (Structured Query Language), where you might find tables with rows (records) and columns (attributes).
- **Unstructured Data**: This type of data doesn't conform to a specific format or structure. It's more free-form, making it more complex to analyze and process.
  - *Example*: The contents of an email, which might include text, images, attachments, or even emojis, or the vast amount of video content uploaded to platforms like YouTube.
- **Semi-structured Data**: Somewhere in between structured and unstructured, this data type might have some organizational properties but doesn't fit neatly into tables like structured data.
  - *Example*: A Twitter tweet, which might have structured data like the timestamp and user ID but also contains unstructured data in the form of the tweet text and any attached media.

Data Collection & Sourcing

- **Public Datasets**: A boon for researchers and developers, these datasets can foster innovation and are often used in academic settings for machine learning projects.
  - *Example*: The MNIST dataset, commonly used in image recognition projects, contains handwritten digits and is publicly available.
- **Purchase or Partnership**: In the data-driven age, data can be a commodity. Companies often buy data or access to it to enhance their services.
  - *Example*: Waze, the navigation app, partners with municipalities to get real-time traffic data to improve route recommendations.

- **User-Generated Data**: Perhaps the most abundant form of data in the social media age, this data is created with every digital action users take.
    - *Example*: Netflix collects data on user viewing habits, which informs its content recommendations and even its content creation decisions.
- **Non-consensual Data Aggregation**: This controversial method involves collecting data without the explicit knowledge or consent of users.
    - *Example*: The Cambridge Analytica scandal, where millions of Facebook users' data was harvested without consent for political advertising purposes. The reason Cambridge Analytica was a huge issue wasn't because of the data aggregation tactics used, but because the data was used for misinformative political campaigning.

      More broadly, AI algorithms are generally trained on data taken from the public domain (e.g., ChatGPT using millions of articles, Wikipedia pages, media, etc. to learn human language)
        - Web Scraping: Beyond just collecting data, web scraping often involves parsing it and storing it in a usable format.
            - Example: Companies like Zillow might scrape real estate sites for property listings to aggregate on their platform.

Processing & Preparing Data

- **Data Cleaning**: Think of this as refining crude oil. It's about removing impurities (errors, duplicates) to get to the valuable product. Without this step, AI models can produce flawed or skewed results.
- **Data Labeling**: In the realm of supervised learning, knowing what each piece of data represents is crucial. It's the difference between showing an AI a picture of a cat and telling it "this is a cat."
    - *Example*: Google's CAPTCHA system, which sometimes asks users to identify objects in images, helps in labeling data for training machine learning models, while simultaneously protecting websites from fraudulent bots.
- **Data Augmentation**: Especially in domains where data is scarce, you can't always collect more data. Instead, you modify existing data to "create" more.
    - *Example*: In medical imaging, where datasets might be limited due to patient privacy, existing images might be modified (rotated, zoomed, cropped) to train models without using new patient data.

Algorithms - How AI is Implemented

If data is the clay for AI, then *algorithms are the tools and techniques* (primarily statistics, differential calculus, and probability) *used to shape the sculpture*. They are the sequence of steps or operations that dictate how data is processed and transformed into meaningful outputs.

Types of Algorithms

- **Supervised Learning Algorithms**: These algorithms are trained using labeled data. The "supervision" consists of the algorithm making predictions and being corrected by the label whenever it's wrong.
    - *Example*: Linear Regression is used to predict a continuous value based on one or more input variables. For instance, predicting house prices based on features like size, location, and number of bedrooms.
    - *Example*: Classification Algorithms like Logistic Regression or Support Vector Machines categorize data points. An email spam filter might use such algorithms to classify emails as "spam" or "not spam" based on their content.
- **Unsupervised Learning Algorithms**: These work on unlabeled data, finding hidden patterns or structures without explicit instruction.

    - *Example*: Clustering Algorithms like K-means group data points based on similarity. Consider how streaming services might cluster songs to create genre-based playlists.
    - *Example*: Dimensionality Reduction Algorithms like Principal Component Analysis (PCA) reduce the number of variables in a dataset while retaining its essential structure. This is useful in visualizing high-dimensional data (data or systems with many different factors that influence them, like diseases) or speeding up other algorithms.
- **Reinforcement Learning Algorithms**: Unlike supervised and unsupervised learning, reinforcement learning is about making a sequence of decisions by interacting with an environment. The algorithm learns by receiving feedback in the form of rewards or penalties.
    - *Example*: Google's DeepMind trained an algorithm called AlphaGo using reinforcement learning to beat human champions in Go, the world's most complex board game.
- **Ensemble Models:** Combining multiple models' decisions to make a more informed final decision. There are multiple types:
    - *Bagging Ensemble Models*: Think of this as asking a group of people the same question separately and then combining their answers. If most of them agree on an answer, it's probably the right one. An example in AI is the Random Forest, which combines many decision trees, and algorithms that follow a defined line of questioning to the data ("if-this-then-this-then-that") to arrive at an output.
    - *Boosting*: Here, imagine teaching a series of students a subject. Each new student learns from the mistakes of the previous student. Over time, you get a set of students whose combined knowledge is vast and comprehensive.
    - *Examples*: Doctors might use ensemble models to diagnose tricky medical conditions. By combining several models' opinions, they increase their chances of getting the diagnosis right. It's a decision-making system reminiscent of the "wisdom of the crowds".
- **Neural Networks & Deep Learning**: These are algorithms inspired by the structure of the human brain, consisting of layers of interconnected nodes or "neurons". The depth (number of layers) and breadth (number of neurons) can vary, leading to a wide range of network architectures.
    - *Example*: Convolutional Neural Networks (CNNs) are designed to process grid-like data, such as images. They excel at tasks like image and video recognition.
    - *Example*: Recurrent Neural Networks (RNNs) are suited for sequential data like time series or natural language. They have "memory" in the sense that their output is influenced by prior inputs in the sequence. A variant of the RNN called the Long Short-Term Memory (LSTM) network has been particularly successful in tasks like language translation or speech recognition.
- **Generative Algorithms (GenAI)**: These are designed to generate new data that resembles a given set of training data. Instead of just classifying or predicting, they can create entirely new content, be it images, text, music, or even videos. GenAI is widely considered the most versatile and general-purpose of the AI panorama. While many of the previously mentioned algorithms can be applied to GenAI, two algorithms are typically exclusively used for generative tasks:
    - *Generative Adversarial Networks (GANs)*
        - Core Principle: GANs consist of two neural networks – a *generator* and a *discriminator* – that are trained simultaneously and work in tandem.
            - Generator: Begins with random input and refines its output over time. Its objective is to produce data so authentic that the discriminator believes it's genuine.
            - Discriminator: Examines data and tries to differentiate between genuine and fake samples. It then informs the generator of how accurate or convincing the generated samples are.

  - Training Dynamics: The training process involves the generator improving its data creation based on feedback from the discriminator. Conversely, the discriminator refines its ability to distinguish real data from fake. This process continues until the generator produces data of such high quality that the discriminator finds it challenging to differentiate from real data.
  - Applications: GANs are versatile and have found uses in creating realistic images, designing video game environments, and even in scientific fields like drug discovery (creating medicines to treat diseases).
- *Variational Autoencoders (VAEs)*
  - Core Principle: VAEs function by compressing data into a more compact form and then reconstructing it. This process ensures that the essential features of the data are retained.
  - Encoding & Decoding: The two primary components of VAEs are the encoder, which compresses the data, and the decoder, which reconstructs the data from this compressed form.
  - Difference from GANs: VAEs differ from GANs in their approach. While GANs rely on the adversarial relationship between the generator and discriminator, VAEs use a probabilistic approach. They add some randomness during the encoding, ensuring diverse outputs. This characteristic makes VAEs suitable for tasks where smooth transitions are essential.
  - Applications: VAEs have been employed in tasks like image denoising, anomaly detection, and even creative endeavors like music generation.
- *Diffusion Models*
  - Core Principle: Diffusion models start with real data like images and gradually add noise. They train a neural network to remove the noise and recover the original.
  - Adding and Removing Noise: First, they take a real image and slowly corrupt it by adding more noise over multiple steps. Then, they train a model to remove all that noise and reconstruct the original image.
  - Difference from GANs: Unlike GANs, diffusion models don't use two competing networks. They simply train on corrupting and restoring real data.
  - Applications: Diffusion models generate realistic new images and videos. They also enhance low-res images to higher resolution.
- *Transformer Decoding Models*
  - Core Principle: Transformer decoders generate text or sequences by predicting one token at a time. Tokens are the basic units that make up the sequence, like words in a sentence. The transformer architecture allows the model to track long-range connections.
  - Generating Token-by-Token: These models are given some initial context. They then predict the next token, and keep generating new tokens step-by-step to build the output. At each step they consider the full context and previous predictions.
  - Difference from Traditional Models: Old models like RNNs process sequences left-to-right. Transformers use attention (a mechanism allows the model to focus on the most relevant parts of the sequence when making predictions) to make connections regardless of position.
  - Applications: Transformer decoders excel at text generation for dialogue, summarization, and creative writing. They also generate music scores and other modalities.

    - *Encoder-Decoder Transformer Models*
        - Core Principle: Encoder-decoders have separate encoder and decoder components. The encoder condenses the input. The decoder uses this condensed version to generate the output sequence.
        - Encoding and Decoding: The encoder summarizes the most important information from the input into a compact representation. The decoder attends to this compact summary and predicts the output sequence.
        - Seq2Seq Models: Encoder-decoders are sequence-to-sequence (seq2seq) models. This means they take a sequence as input, process and condense it, and generate a new sequence as output.
        - Uses: Encoder-decoders have been very successful for machine translation, converting text between languages. They also summarize text and answer questions.
    - *Examples* of generative models include **OpenAI's ChatGPT** chatbot (a seq2seq transformer model) and **Anthropic's Claude** chabot (which combines deep learning and reinforcement learning to create a safe, "constitutional" language processing and interaction algorithm), and **OpenAI's DALL-E**, a GAN capable of generating or editing images using text and image prompts.

Learning in Algorithms

The central thesis of machine learning, the most common and relevant area of AI, is teaching machines to learn from data as the name denotes. This is accomplished through a series of steps.

- **Initialization**: Before AI starts its learning process, we set some initial values to its internal parameters. These are akin to the initial settings or adjustments we make before starting any new task. While these initial values might not be perfect, they give the AI a place to start.
- **Feedforward**: This is the first step of the AI's decision-making process. The AI takes in data, processes it layer by layer, and produces an initial output or prediction. Each layer refines the data, making it closer to the desired output.
- **Loss Calculation**: Here, the AI evaluates its performance. It compares the prediction it made in the feedforward step to the actual, known outcome. This comparison produces a "loss" value, which simply quantifies how off the mark the AI's prediction was.
- **Backpropagation**: Based on the loss value, the AI goes back through its internal processes to adjust its parameters. This step ensures that the next time the AI encounters similar data, it can make a more accurate prediction. It's a feedback mechanism: the AI identifies where it went wrong and tweaks its approach to do better next time.
- **Iteration**: The entire process – from feedforward to backpropagation – is repeated multiple times. With each iteration, the AI improves its internal parameters, refining its predictions. The more repetitions (or iterations) the AI undergoes, the better it becomes at its task.

AI's vast adaptability is both its greatest asset and liability to the general public. While its broad capabilities promise transformative solutions, they also introduce significant risks, especially when their applications are unbounded. As we navigate this AI-driven era, understanding its technical essence is crucial to both establishing standards of development (which will help ease market fragmentation in the space and encourage antitrust/broader development), informing what aspects of it should be regulated, and determining how to categorize risk. Furthermore, properly defining AI is foundational to crafting regulations that safeguard without stifling innovation and align with the blueprint for the AI Bill of Rights.