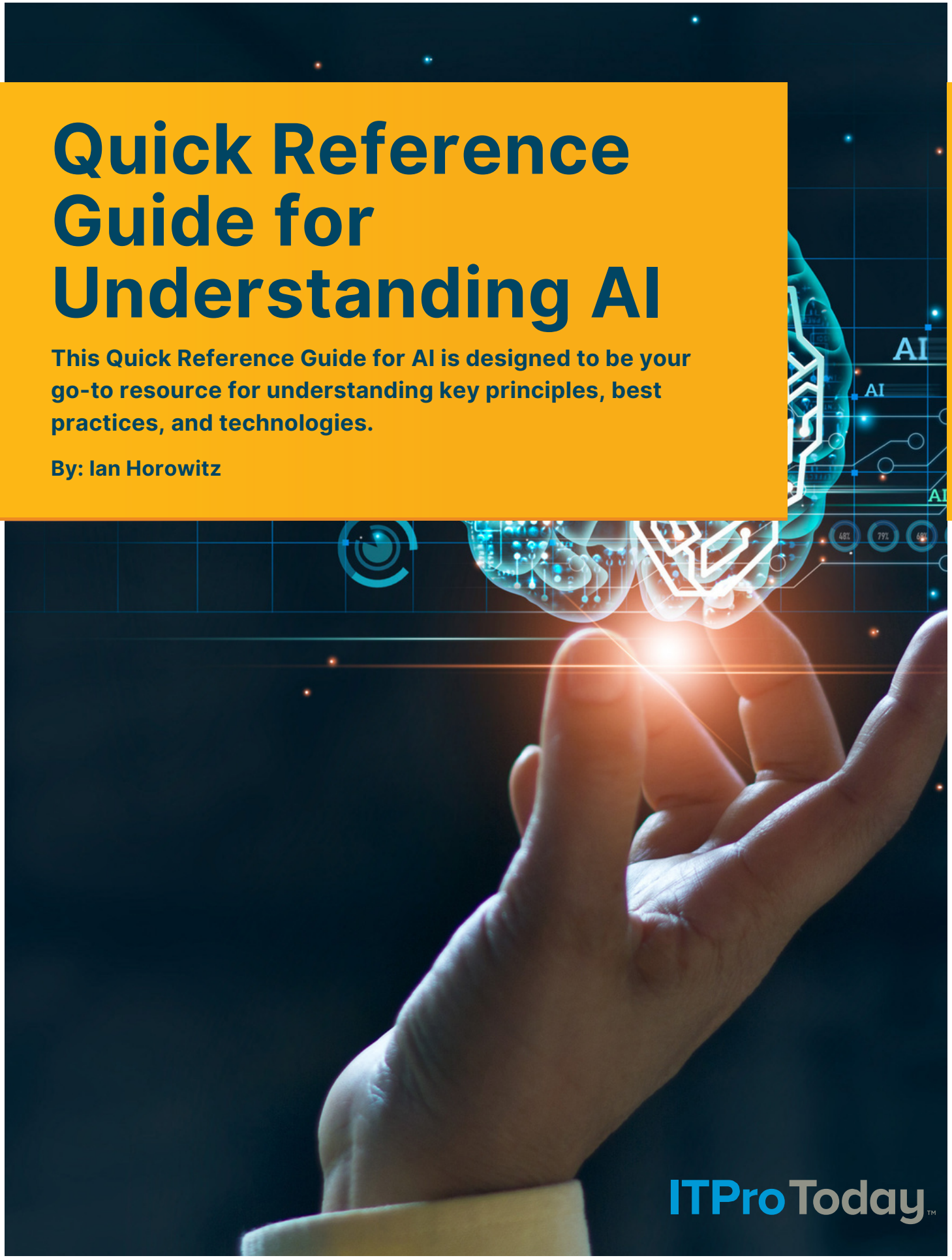
Quick Reference Guide for Understanding AI

This Quick Reference Guide for AI is designed to be your go-to resource for understanding key principles, best practices, and technologies.

By: Ian Horowitz



Introdution

Artificial intelligence (AI) is the superstar of the tech world today, with virtually every industry considering how AI can enhance products, services, and competitive strengths. At the same time, discussions about the ethical use of AI have led to persistent concerns and the creation of new policies and laws. While initial impressions of AI may evoke a mix of excitement, skepticism, worry, and countless questions, it's important to recognize why AI matters. AI aims to replicate human-like cognitive abilities through computational models and underlying algorithms. By providing machines with the power to learn, adapt, and perform tasks that usually require human intellect, AI represents a transformative era in computing and IT.

This Quick Reference Guide is designed to assist you in exploring fundamental AI principles, best practices, and cutting-edge technological developments. Inside, you'll find essential AI terminology and clear explainers to help you start your AI journey.

Table of Contents

1. Types of AI

Narrow vs. General AI

When we discuss AI, we refer to two overarching types of systems: Narrow AI and General AI.

Narrow AI (also known as "Weak AI") is designed and trained for specific tasks. It operates within a limited context and awareness, which is typical of most AI systems today. Narrow AI excels in executing singular tasks such as facial recognition or language translation but can't extend its intelligence beyond its programmed scope. Examples include AI-powered search engines, recommendation algorithms, chatbots, and voice assistants like Apple's Siri and Amazon's Alexa.

In contrast, General AI (also referred to as "Strong AI") is a theoretical concept of intelligent systems that could potentially understand, learn, and apply machine intelligence across a broad spectrum of tasks, displaying human-like or even *transhuman* cognitive abilities. In theory, General AI could possess or demonstrate some level of sentience, self-awareness, emotion, and an understanding of abstract concepts.

Machine Learning (ML)

Machine Learning (ML) is a subset of AI based on enabling machines to do two things:

1. Learn from inputs.
2. Use pattern recognition to make complex inferences and decisions.

ML operates on underlying algorithms that get better over time through learned experiences and training data.

ML is broadly classified into three main types:

1. **Supervised Learning**: The algorithm is shown a lot of examples (i.e., labeled datasets) of how it should "think" until it can make accurate predictions. As the algorithm recognizes patterns and relationships in the data it's shown, it adjusts its internal parameters to make its predictions increasingly accurate. When the algorithm makes mistakes, it can be given feedback to improve its generalizations.
2. **Unsupervised Learning**: The algorithm is given a lot of unlabeled data and made to identify patterns, connections, or structures within the data without any clear external guidance. Unsupervised learning is commonly used for vast datasets because it can discover relationships that would otherwise be overlooked through manual assessments.
3. **Reinforcement Learning**: An algorithm or entity (called an "agent") learns by receiving rewards or penalties as it explores an environment. The agent uses trial and error and is given a reward signal or penalty signal based on its actions. The agent then adjusts its strategy based on this feedback.

ML has a variety of applications.

Examples:

- Recommendation engines used by companies including Amazon, Meta, and Netflix.
- Records processing and diagnostic assistance in AI medical tools, such as those spun off by IBM's failed Watson health initiative.

Deep Learning

Deep learning (DL) is a subfield of machine learning. Deep learning relies on multi-layer neural networks, including Artificial Neural Networks (ANNs),

Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs). Neural networks enable machines to process data in nuanced, almost human-like ways.

DL models excel at identifying and interpreting elements from basic, unprocessed data. This makes them adept at performing complex tasks like image and speech recognition.

A quintessential application of deep learning today is in the field of computer vision (CV), where it enables machines to interpret and understand the content of images and videos. This technology plays a key role in the ability of autonomous vehicles to "see" and navigate safely. Current production models of Tesla vehicles use neural networks, which are largely reliant on GPUs from AI hardware and software vendor Nvidia.

Natural Language Processing (NLP)

Natural language processing (NLP) is a part of machine learning that sometimes incorporates deep learning models. NLP focuses on the interaction between computers and human language, teaching machines to understand, interpret, and respond to language.

A primary application of NLP is language translation services, such as those seen in Google Cloud's translation tools. In addition, NLP, along with deep Learning, has been important in the development of sophisticated Large Language Models (see the next section). These models are behind popular chat-based tools such as OpenAI's ChatGPT and Google's Gemini.

Large Language Model (LLM)

Large Language Models (LLMs) are AI systems designed to understand, generate, and interact using human language. These models are trained on vast datasets of text, enabling them to grasp language patterns, nuances, and contexts. Text generation models, like those developed by OpenAI, generate text based on the input they receive.

The applications of LLMs range from automating email responses and organizational office tasks to more complex tasks such as writing code.

Generative AI

Generative AI, or GenAI, is a type of AI system that uses algorithms to generate content. GenAI systems today can produce content in a variety of mediums, including the following:

- **Text** – prose or verse in basically any format and genre, including nonfiction and fiction.
- **Pictures** – photorealistic and creative imagery in imitations of artistic styles.
- **Video** – videos and animations.
- **Audio** – imitations of human voices as well as music composition.

Microsoft's Copilot is an example of a system that uses GenAI and LLMs to generate personalized assistance to users. For example, Copilot is integrated with various Microsoft 365 apps, such as Word and Excel, to help with everyday work tasks.

GenAI can also be used to generate data. For example, GenAI can create synthetic data to help fill out datasets for use in machine learning models.

Many fields are currently using GenAI, such as pharmaceuticals, where it can help create molecular structures for new drugs.

You will learn much more about GenAI throughout this guide.

2. How Is AI Used Today?

Let's look at how several major industries are applying and testing AI technologies.

Financial Services

In the banking and financial services sector, AI has become instrumental in detecting potential fraud and suspicious activity. Machine learning algorithms, in particular, can be useful in fraud prevention. Companies like PinDrop use machine learning for user verification and fraud prevention in the banking sector.

Healthcare

Despite IBM's AI incursions into the healthcare industry not paying off, AI is making a significant impact in drug discovery and diagnostics. AI can expedite the drug development process by analyzing large datasets, predicting molecular interactions, and simulating drug responses. This allows for faster identification of potential drug candidates and the development of new therapies.

From big tech companies to smaller startups, AI-driven drug development is widely seen as the catalyst for a coming revolution in healthcare. Interest is reflected in substantial venture capital funding, reaching $2.1 billion in the first half of 2021 alone. Google's DeepMind subsidiary has also been reported to have made large investments in this type of technology.

Retail

AI has a growing role in delivering personalized shopping recommendations. Various retailers and online marketplaces, most notably Amazon, rely on machine learning algorithms to push recommendations to their customers.

Recommendations powered by AWS' Amazon Personalize service are tailored to users' preferences based on their interactions, purchase history, and browsing behavior. Amazon Personalize gives retailers and service providers access to the same ML algorithm used by Amazon.com's recommendation system.

Although not "retail" businesses in a traditional sense, subscription-based media and streaming services, including Spotify, rely heavily on machine learning. They deploy sophisticated recommendation algorithms to improve subscriber engagement and retention by anticipating users' interests.

Manufacturing and Industrial

In manufacturing, AI has proven vital in improving efficiency and reducing downtime. In addition to automation, a common AI application is predictive equipment maintenance. Machine learning algorithms enable factories to predict when equipment might fail or need maintenance.

Major industry players, such as General Electric, Honeywell, Siemens, and IBM, offer proprietary versions of AI-powered technology for predictive maintenance. The space is also filled with smaller competitors offering custom products, ranging from IoT monitoring and service log maintenance to AI-based warning systems.

3. The Field of AI Ethics

Responsible AI

Incorporating ethical principles into the design and development of AI systems is necessary to mitigate biased decisions. In discussions about Responsible AI, an AI framework that focuses on ethical and legal interests, you will come across concepts and principles such as the ones we define here (in very simple terms):

- **Data bias** – The datasets used to train AI models are biased and create an incomplete picture because they lack diverse and representative information.
- **Inherent data bias** – Bias is ingrained into the datasets used to train AI models due to the nature of the dataset itself.
- **Fairness** – In an AI context, fairness is the assurance that AI systems don't favor one group over another and are free from biases in their decision-making processes.
- **Transparency** – Transparent AI systems means that they are open to scrutiny and capable of being explained by AI developers.
- **Explainable AI (XAI)** – Explainable AI describes AI systems that can demonstrate how they arrive at a decision. In contrast, "black box" AI systems can't show how they make decisions.
- **Accountability** – In an AI context, accountability means AI developers and users take responsibility for the impact of an AI system and its decisions.

Regulatory frameworks and guidelines for AI deployment are currently in development. The European Commission has proposed a legal framework for AI, which includes clear requirements for high-risk AI systems, suggested market regulations, and stringent enforcement measures.

In the U.S., although there is no comprehensive federal legislation governing the sale, marketing, or use of AI, various initiatives and guidelines encourage the adoption of explainable AI over black-box AI models.

Deep learning is frequently described as a black-box AI model because it tends to lack interpretability and explainability. In fact, complex deep learning and neural network architectures can be so opaque that computer scientists and researchers may struggle to understand how their systems work and make decisions.

The National Institute of Standards and Technology (NIST) has developed a set of principles for explainable AI, aiming to address concerns related to interpretability and explainability deficits seen in some AI systems.

Fairness and Bias

AI is increasingly handling complex decision-making tasks. As such, bias must be minimized to ensure fairness and equity. However, weeding out bias in AI represents a huge challenge. As AI systems grow more intricate, interpretability becomes more difficult. This highlights the need for maintaining accountability through the responsible design and deployment of AI systems.

Privacy

Many people inside and outside of the tech industry agree that it's incredibly important for AI systems to respect the fundamental rights, privacy, and autonomy of users. To ensure this, the development and use of AI should be guided by stringent safety measures, incorporating human-based monitoring and auditing processes to reduce potential risks.

To monitor AI technologies for privacy breaches, organizations must implement testing procedures, external validation, and ongoing oversight. This becomes critical when operating in markets outside the U.S., where data protections are stronger – e.g., in the European Union, which has the General Data Protect Regulation (GDPR) laws.

GDPR governs data use and offers individuals significant control over personal data. These protections affect how AI systems can be designed and used in the EU. Companies deploying AI systems internationally must also be aware that

individual countries have their own data privacy policies. For example, Italy temporarily blocked access to ChatGPT in 2023 due to its data privacy policies

4. AI Frameworks

TensorFlow

Developed by Google, TensorFlow is an open-source framework widely used by AI developers for its flexibility and comprehensive toolkit.

TensorFlow is well-suited for large-scale, complex computations and offers extensive support for neural networks. Many deep learning applications are built with the TensorFlow framework.

TensorFlow is popular in both research and production settings, excelling in areas such as speech and image recognition and predictive analytics.

PyTorch

PyTorch, created by Facebook's AI Research lab, has gained popularity for its user-friendly interface and dynamic computational graph. It's particularly favored in the academic and research community due to its straightforward way of creating and experimenting with neural networks. PyTorch’s flexibility allows developers to make on-the-fly changes to complex network architectures.

PyTorch is part of the Linux Foundation’s collection of open-source software for developers.

scikit-learn

scikit-learn is mainly used for traditional machine learning algorithms. Built on top of SciPy (Scientific Python), its data mining and analysis tools make it a popular choice for ML-dependent applications. Applications include processing large datasets, automating industrial processes, and other routine activities where a high level of explainability is desired.

5. AI Security

AI needs data for training, validation, and operation. As such, data protection a top priority.

Similar to any server or cloud-based technology, AI systems can have vulnerabilities that leave it exposed to cyberattacks and data breaches. In addition to safeguarding user data, IT professionals must vigilantly protect AI systems from malicious manipulation, which could lead to harmful outcomes. For example, subtle changes in input data by internal or external actors can lead an AI system to produce erroneous or biased outputs.

The security of AI systems requires a multi-layered approach, integrating advanced encryption and secure data storage technologies. It also requires stringent multifactor authorizations and access protocols.

6. Regulations and Compliance

While we've already discussed the GDPR and the EU's proposed legislation on AI, for IT professionals operating in the U.S. market, it's a good idea to familiarize yourself with the California Consumer Privacy Act (CCPA).

The CCPA deals primarily with consumer data rights. It ensures certain protections for how data can be handled and sold, also specifying that companies must disclose how they use data. For any AI system that processes large amounts of U.S. consumer data, it's important to fully comply with with the CCPA. Additional U.S. regulations are likely ahead.

About ITPro Today

ITProToday.com is a leading online source of daily news, analysis, opinions and how-to's about the information technology industry. Along with offering practical IT operations and career insights, we help IT professionals and technology stakeholders learn about, assess and manage the acquisition of next-gen technology that drives business innovation, including — but not limited to — analytics, artificial intelligence/machine learning, cloud computing, low-code/no-code, DevOps, NoOps, DataOps, compute engines, containers, edge computing, hyper-converged infrastructure, security, software development and storage.