CHAPTER ONE

1 Introduction

Summary

After reading this chapter, you will be able to:

- Understand who this book is for.
- Understand the basic building blocks of a neural network.
- Understand the current relevant regulatory frameworks and their impact in deep learning in model development.

The world is complex and multifaceted. This is not a controversial opinion. In fact, nuance and care in the development of models is paramount in banking. For decades, borrower behavior has been modeled and managed through data science models of increasing complexity. In credit scoring, for example, we have moved from tree-based models in the 1970s to logistic regression dominating the landscape until very recently. Currently, however, there has been a shift toward tree ensembles such as XGBoost and Random Forests in jurisdictions that allow it. This last shift is relevant as it shows how banking, while slow to adapt, can move toward more advanced models when the profits are clear and the regulatory hurdles can be overcome.

This book is our vision on what we believe is the next evolution of banking: the use of alternative data, that is, data that go beyond simple tabular variables, into new models that can consider complex inputs in a transparent and explainable manner. The pages in this book discuss this in detail, trying to always solve the balancing act that we believe makes artificial intelligence (AI) in banking so interesting. It must arise from a specific business need and show that it is a profitable endeavor, while at the same time it needs to balance the rights of individuals as reflected in the specific regulation that is pertinent to its customers. Solving this puzzle requires creativity, strong intuition for data capabilities, computational skills, computational resources, regulatory acumen, and business sense. There are very few areas where data science is applied that require all of these skills to work together simultaneously.

1.1 WHY WE WROTE THIS BOOK AND WHO IT IS FOR

Answering the first question: we wrote a book we wanted to have on our desks and to point our business partners, students, and interested parties to. Books on deep learning are easy to find, but comprehensive texts on banking regulation are far scarcer. Separately, while numerous books address topics like credit risk and model risk, they treat them as distinct subjects. What has been missing is a unified guide that combines all three areas. The book discusses, through its nine chapters, alternative unstructured data, where such data appear in banking, the regulation that applies to it, the techniques that are relevant to tackle the problem of modeling these data using deep learning, and case studies with real data applying everything we have discussed. Our intention is that the book points you, the reader, in the right direction and guides you as you first identify a source of data that can help you solve a problem, then develop your own models to tackle the problem, and evaluate them using both quantitative and qualitative measures common in banking. Chapter 9 also discusses briefly how to deploy these models in practice and where future regulation may go.

Answering the second question: this is a technical book. We do not shy away from mathematical formulas, from computational code and algorithms, from regulatory discussions, nor from complex descriptions of advanced models and how they are applied. Although we try our best to keep the language simple and direct, this is a book intended for a mature audience with some knowledge in the area. We expect the readers to have training in data science, data management, and model development and deployment. We start by assuming that the reader has some knowledge in these areas, although we do not assume deep learning knowledge. From a data science point of view, we assume that

the reader understands how to train and deploy an XGBoosting model and is familiar with neural networks, to give a slight benchmark, at a level consistent with an undergraduate course in data science. The book will guide you to gain the necessary knowledge (through the methodological sections) and skills (through the case studies) to become effective at deploying deep learning models on top of your traditional modeling skills.

From a regulatory point of view, we expected the reader to understand the frameworks that are relevant to the practice of banking. We focus mainly on risk modeling regulations (Basel & IFRS 9, AML/CFT accords, and local regulations in selected territories), AI regulations (AI Act in Europe and similar proposals worldwide), and data privacy regulations (EU GDPR and comparable legislation). If you are unfamiliar with these, we do try to explain further what parts of these legislations, particularly for AI and consumer privacy, apply for deep learning, although some background knowledge is required of the banking-specific ones.

If you are an academic and are wondering whether this book would help your courses, we intend this book for either a follow-up to a data science algorithms course, when the new course focus is on applied deep learning with an orientation in banking, and also for banking analytics or applied deep learning courses at a graduate level. This is a book in banking and deep learning, and we believe courses that provide these specific skills, or general applied deep learning courses that are looking for complementary texts in a specific area, would be well-served by this textbook and labs.

Now that you have decided that this book is for you, the next question you need to answer is whether your problem actually needs deep learning.

1.2 DO YOU NEED DEEP LEARNING?

Let us start by thinking about whether there is even a need to deploy deep learning. There are three questions that a savvy manager can ask themselves:

1. Is there a problem that we have not been able to solve with traditional models?
2. Do we have data in our data lakes/data warehouses that is not being used or is underutilized?
3. What is our current technical capacity?

Within different organizations, there will most likely be a mix of answers to these questions. Let us imagine a traditional bank that has been collecting the

social media mentions of their Small and Medium-sized Enterprise (SME) customers. This information is used in their marketing propensity models by simply counting the number of mentions in a 90-day period to identify whether the SME is generating buzz and may have a need for funds the bank can provide. This information is combined with financial transactions, and a simple regression model generates a propensity probability. Starting with the first question, the answer may be that the model fails to identify negative buzz, and thus generates incorrect offers to companies that are in a downturn or subject to media controversy. The second question follows, and the answer is that we are indeed underutilizing the text data in the data lake, as we are only generating this buzz indicator, without considering the context that a more sophisticated model can bring. The third question is much more complex to answer. If the managers of our example institution desire to move forward with a more complex multimodal model, they would need to identify if they have the correct collaborators who can develop the model in their data science teams, and whether there is on-premise or cloud infrastructure that can be used to develop the models. Cloud Graphics Processing Unit (GPU) infrastructure can get expensive fast, and on-premise infrastructure may not be sufficient to train a model only to deploy them (inference).

One of our own coauthors recently published a paper arguing against the use of deep learning for tabular data (Gunnarsson et al., 2021). If you are simply expecting that plugging structured, tabular, data into a fancy deep learning model will lead to immediate gains, we are sorry to tell you that it probably will not. Most likely, this will increase your inference costs hundredfold and will provide, at best, a marginal gain. Moving from a logistic regression-based model to a tree-based ensemble is probably a much better use of resources, as was shown a decade ago by our good friends in Lessmann et al. (2015). Deep learning shines in three aspects: when your data is unstructured, when you need to combine different sources of data, and when you have complex data structures that you need to use, such as time series of different or unknown lengths. Otherwise, traditional models will serve you best.

If your data is beyond tabular (which it probably is), then the question of using deep learning is much more interesting. This book is for those cases. Let us consider a few examples where this book may help.

1. When evaluating a mortgage, structured data can be misleading. For example, a specific community may have invested much in better greenery, more trees, a nice park, or any other kind of improvements that can make a specific valuation of a property significantly out-value neighboring

properties. What alternatives are there to evaluate these intangibles? In Chapter 2, we show the case of evaluating mortgages using Light Detection and Ranging (LiDAR) data to assess environmental characteristics and value properties more effectively, complementing traditional data.

2. In Tavakoli et al. (2025), we showed that, in times of crisis (in particular COVID-19), the predictive power of earning calls over future rating changes. Effective managers can provide context that would otherwise be missing, and such context is, in turn, a good predictor of future performance. In Chapter 4, we analyze how the Fed hints at future economic performance, following a similar logic.
3. An important and often ignored fact about credit risk is that it is correlated. This correlation arises because risk is intrinsically a shared event. A company closing its doors means that all its former workers are at increased risk of default. A plague in a region can affect all farmers in an area, leading to a higher risk. Unfortunately, an ever more common natural disaster leads to an increased default rate in the area it affects. Such contagion risks, spillover risks, or correlated default risks, as they are known in the literature, are more often than not neglected. But networks are commonly available in banking, they arise from transfers, purchase orders, factoring, and can be inferred by geographical location and economic sectors, among other variables (Oskarsdottir and Bravo, 2021). Chapter 5 discusses how to utilize them in your models.
4. What if you do not necessarily have large amounts of unstructured data but different types of data? Chapter 3 deals with the particular case of time series, where complex mixed time series can be combined into powerful models. In Chapter 7, we go beyond this and discuss the case of multimodality, where we can combine models from previous chapters to construct complex models with multiple branches, allowing a user to, for example, identify the risk of a mortgage borrower, including both how their network influences potential contagion and the risk arising from the property valuation deduced from the LiDAR image of their neighborhood.

These are just some examples that we will cover in this book. During our research, we have found increasingly more evidence of the value of alternative data for credit risk. Regulators around the world have also accepted the idea that credit risk models can be enriched by deep learning, as we will discuss in the next section. In addition, there are many areas where unstructured data is the main source of data available, such as text-based suspicious activity reports, common in anti-money laundering activities.

1.3 BOOK STRUCTURE

Our aim in this book is for it to be a one-stop reference in the use of alternative data in banking, particularly in retail risk areas, but we hope users in nontrading areas also see value in our work. As such, we divide the books into how specific types of data can be used. The chapters are all structured into a type of data and a type of models. For example, the next chapter is about image data and how to use convolutional neural networks to leverage them.

Within each chapter, we start by introducing the data and why it is relevant in banking. We give an overview of where these data may arise in banking practice. The purpose of this section is to help the readers consider whether they have these types of data available somewhere else on their data lakes. A question we often ask our partners is, "What data do you not use now?" Maybe in that section you will remember that there is a source you had missed and your models will be richer for it.

Next, and if applicable, we discuss any specific regulatory issues regarding that data source. In this chapter, we give overview of regulations that intersect regarding banking overall, but for many specific types of data, there may be unique indications and considerations that a modeler needs to take into account. An example is the mandate to use text data arising from suspicious activities reports (SAR) that we show in Chapter 4, or the chance of biasing models when using images that we discuss in Chapter 2 and then provide the tools to identify in Chapter 8.

Next, we go in-depth on the techniques and methods that underpin modern models that can use those data sources effectively. We have aimed to be extensive here and not shy away from the math. Some readers may find these sections too extensive. We, again, are aiming to be a one-stop book. Although we acknowledge that it is impossible to cover every little detail that is possible to cover in AI models, we have attempted to give as much of an overview of methods as possible.

We finish every chapter with a case study of applying a subset of these models to a banking problem with real data. We are fortunate to have been able to secure legal permission to use Freddie Mac's data for this work, which forms the basis for many of the case studies in this work. In particular, we use the Single Family Loan Dataset,[1] which deals with single family mortgages that have been reinsured by Freddie Mac. It is a rich source of time series data and a base to connect them with other freely available sources of unstructured data, such as LiDAR images. We also use speeches from the Fed and the

[1] https://www.freddiemac.com/research/datasets/sf-loanlevel-dataset

Basel Committee on Banking Supervision (BCBS), market data, and macroeconomic data from public sources to give a wider range of cases in the book. Every chapter, except this one and the last one, focused on a future-looking discussion of the evolution of AI in banking.

The case studies have a particularity that we hope increases the shelf life of this book. We describe the algorithms in language-agnostic pseudocode form. This does not mean that there is no code for this book. You can find the implementations of the algorithms on the book's website.[2] These are implemented on a GitHub site that we would also love to grow into a live site that reflects the best practices in deep learning for banking. Have you found a more effective parameter combination? Optimized something? Feel free to make a pull request! We will also keep these labs updated with the latest changes in best practices, without significantly altering its structure from what the book's pseudocode shows. We have made the labs run on what is currently available for free in the cloud. In particular, all labs, except for Chapter 6's, can be run on Google Colab's T4 GPU that are available for free.

Now that we have introduced the book, for whom it is for and its structure, we can start with the content itself. In the following section, we will discuss how AI and banking regulations intersect. We close this chapter by providing the basis of modern neural networks, before moving into our first data type and models to use them: images and convolutional neural networks in Chapter 2.

1.4 SOME RELEVANT REGULATORY FRAMEWORKS

Unlike the relatively uniform evolution of banking regulation under the Basel Accords, the emerging regulatory responses to AI in credit risk have been fragmented and jurisdiction-specific. Although most regulators recognize credit risk as a high-impact area of AI application, their approaches to governance, fairness, data usage, and model oversight vary significantly in both scope and implementation. Together with coauthors in Germany and the United Kingdom, we recently published a discussion paper on the topic of divergence of AI regulation, and the traditional Basel accords (Bravo et al., 2023). This section is based on this work and updates the regulation with the latest developments.

The regulatory divergence referenced previously is driven in part by differences in legal traditions, institutional capacities, and policy priorities. For example, some jurisdictions emphasize AI ethics and data governance, while others focus narrowly on model explainability or consumer protection. In addition, regulatory responses have emerged in a bottom-up fashion, often led by

[2] https://bankingbook.ml

national legislative or supervisory bodies, rather than through a coordinated top-down process similar to that of the Basel Committee.

Most regulators articulate the need for explainability, fairness, human oversight, and sound model governance, reflecting a convergence around certain foundational principles. However, this convergence remains conceptual: few concrete regulatory instruments provide consistent definitions, enforceable standards, or interoperable methodologies across borders, but there is a wide-spread understanding of this need.

Lack of regulatory alignment carries material risk. Divergent standards on fairness or permissible data use—particularly with respect to unstructured or alternative data—could undermine trust, distort capital allocation, and exacerbate inequalities in credit access.

We believe that as national AI laws begin to intersect with long-established banking regulations, there is an urgent need for convergence in technical requirements, interpretative guidance, and supervisory expectations. A harmonized global framework would not only support financial stability and regulatory clarity but would also promote innovation and equitable access to credit. We, of course, also recognize the gigantic challenge such a regulation would entail. In what follows, we discuss specific regulatory environments and their characteristics for some of the most advanced territories in AI evaluation. With our apologies to our readers outside of these territories, we have focused on the United States, United Kingdom, Canada, and the European Union, as we did in our original paper. We also added Singapore, as they have made significant efforts to close this gap.

1.4.1 The European Union

Without a doubt, the European Union is the territory that is further along in regulation of AI, to the point that there is a clear regulation in force across it, with published guidelines by the corresponding regulator, the European Banking Authority.

The EU AI Act (Reg. 2024/1689), adopted in June 2024 and in force from August 2024, explicitly covers credit scoring (European Parliament and Council of the European Union, 2024). Under the Act, any AI system used to assess credit scores or creditworthiness of natural persons is classified as high-risk. High-risk systems must meet strict obligations: risk management and impact assessments, use of high-quality (nonbiased) data, robust monitoring and logging, transparency to users, and human oversight. In practice, this means EU banks that use AI in lending must implement governance controls, bias checks, and documentation far beyond previous norms. However, the

Act exempts pure fraud detection systems and internal capital modeling from the high-risk label, focusing on customer-facing lending uses. It also explicitly focuses on natural persons, leaving corporate and SME as a rich opportunity for the use of advanced methods. Of course, existing Capital Requirements Regulation and Credit Requirements Directive (CRR/CRD) rules on capital and risk management also remain enforced.

On the banking regulation side, the reforms in 2022–2024, such as CRR3/CRD6 implementing the Basel III endgame, update capital and risk rules but do not explicitly mention AI. The main intersection with the AI Act is through overlapping requirements: for example, EU banks under CRR must validate internal credit scoring models—now those models would also fall under the AI Act if they use AI techniques for consumer credit. In other words, the AI Act requirements are an overlay on top of existing Basel-based rules. The European Banking Authority has also monitored AI: its Risk Assessment notes from December 2022 (European Banking Authority, 2022) cite widespread use of AI in credit scoring and risk models, but its guidance remains largely based on principles. Interestingly, the EBA has explicitly mentioned too that nonlinear AI models can be used for credit risk, as long as the bank can prove that those models bring significant gains in risk management with limited risks. Furthermore, the far-reaching General Data Protection Regulation of 2016 (The European Parliament and the Council of the European Union, 2016) had already provided significant protections to EU residents regarding their personal data and privacy rights. The AI Act adds further layers of regulation on AI usage and how usage of data for the specific purpose of AI models should be treated. This adds, without a doubt, a layer of complexity to any AI developments in Europe, as companies need to navigate these multiple legislative guidelines.

The AI Act includes a series of recitals that provide some exceptions to AI use in high-risk environments. Recital 53 suggests that an AI system might not be considered high-risk, even if it is used in a high-risk area, if it meets a set of cumulative conditions. One of these conditions is that the system is intended to perform a "narrow procedural task." The recital gives the specific example of a system that "transforms unstructured data into structured data." Other conditions include that the system is intended to improve the result of a previously completed human activity or is purely preparatory to a final assessment. For example, an AI that simply improves the grammar of a previously drafted document or translates a document would likely fall under this exemption. However, the recital concludes with a critical caveat: "Any AI system that engages in 'profiling' as defined under the GDPR will always be considered high risk if it is used for one of the purposes listed in Annex III, such as credit scoring."

In unstructured/alternative data specifically, the AI Act demand for "high quality training data" implicitly restricts the use of noisy unvetted sources in consumer credit, though the Act does not ban them outright. Overall, the EU approach augments Basel principles: like Basel, it insists on risk governance, but it adds new explicit fairness/transparency duties for AI in lending and wide-ranging data protections via the GDPR. Compared to other jurisdictions, the EU is uniquely prescriptive (mandatory compliance requirements), whereas Basel and most countries had relied on broad model-risk frameworks.

1.4.2 The United States

The United States has a significantly divergent regulatory landscape, even within its own borders. Some states have enacted ambitious regulations regarding the use of AI, such as California and New York. However, at the time of writing in 2025, the United States is rolling back many of these and forbidding the enactment of new ones. A provision in the colorfully named "One Big Beautiful Bill" prevents states from regulating AI for 10 years unless they return broadband development funds to the federal government (United States Congress, 2025).

The Consumer Financial Protection Bureau clarified that the Equal Credit Opportunity Act (ECOA) adverse action notice rules apply fully to AI/algorithmic lending (Consumer Financial Protection Bureau, 2023b). Lenders cannot rely on generic checklists when explaining credit denials; they must give consumers specific and accurate reasons even if a "black box" model was used. This guidance (Consumer Financial Protection Bureau Circular 2023-03) reinforces the ability to explain and prevents lenders from hiding behind the complexity of AI.

The Federal Reserve and other regulators have issued a landmark interagency statement underscoring the governance and auditability of all automated models (Board of Governors of the Federal Reserve System et al., 2023). This statement explicitly mentions the allowable use of *second-look models*, or the use of a model, which can be more complex or uses alternative data, to take a secondary decision if the regular, more transparent model decides against a lending decision. These strategies can help with financial inclusion while also allowing one to show the value of AI models.

In general, these regulators emphasize a technology-neutral, risk-based framework. Their position is that AI tools should follow the same governance as any model. In June 2024, the Acting Comptroller of the Currency stated that banks must apply traditional model-risk practices to AI systems and called for transparency and accountability in AI use. The Office of the Comptroller of the

Currency (OCC) guidance has confirmed the interagency statement on alternative data to enhance credit models for financial inclusion, while still expecting the same risk controls of the Office of the Comptroller of the Currency (2024).

In summary, US banks using AI for credit must already comply with ECOA, the Fair Credit Reporting Act, and antidiscrimination laws. The Basel-aligned capital rules (US Basel III final rules) do not specifically address AI models—instead, fair lending and model validation requirements remain in place. The United States approach thus converges with Basel principles (sound model governance, explainability, and fairness) but does so through general laws rather than a dedicated AI statute. AI-driven credit systems must still meet oversight, and regulators have signaled that they will not accept opaque AI-driven denial practices. However, the federal government seems, at this time, much more open to the use of other models, so the evolution of the United States is uncertain.

1.4.3 The United Kingdom

The United Kingdom is another territory that has enacted a stand-alone AI law to date, favoring a sectoral, principles-based approach. The Acting Comptroller of the Currency (Michael J. Hsu, 2024) explicitly state that the United Kingdom "does not intend to enact horizontal AI regulation in the near future" (UK Government, 2023) rather than leaving AI oversight to existing regulators under common principles. In practice, this means credit scoring models must comply with current financial and consumer laws (FCA Principles, Consumer Duty, Equality Act, etc.) rather than any new AI-specific statute.

The Financial Conduct Authority (FCA) and the Prudential Regulation Authority (PRA) have issued guidance stressing robust model governance. For example, in Financial Conduct Authority (2025b), they tested different ways of explaining AI-driven credit decisions to consumers and found that the explanation style greatly affected the ability of users to detect errors. The Bank of England/PRA guidance (e.g., the CP6/22 rules and its supervisory statement SS1/23 in 2023) reinforces traditional model risk management (documentation, validation, and oversight) for all models, implicitly covering AI (Prudential Regulation Authority, 2023).

The United Kingdom has not introduced special rules on unstructured or nontraditional data in credit models. Any use of alternative data must still respect data protection and anti-discrimination laws. They continue to apply familiar principles, such as the PRIN 6 "Customer Interests," and the discrimination ban of the equality act, to algorithmic models. Although the UK's Consumer Duty does not mention AI by name, it requires fair outcomes and

governance that would capture AI-driven credit models (Financial Conduct Authority, 2022). The PRA model risk principles similarly call for independent model validation and senior oversight of all models (Prudential Regulation Authority, 2023). These reflect Basel-aligned expectations (sound governance, testing, and documentation) rather than novel rules.

In general, the UK strategy converges with Basel-era principles (sound risk and governance) but diverges from the EU: unlike the EU AI Act, the United Kingdom does not classify AI scoring credit as "high-risk" with new mandatory requirements. We expect the newly reintroduced AI Act/Bill and future regulator initiatives (e.g. the FCA's AI Strategy) to layer on existing financial regulations, not overhaul them.

1.4.4 Canada

Canada has no AI-specific law in force yet. A federal AI Bill (the "Artificial Intelligence and Data Act," Bill C 27) was tabled in 2022 but did not pass before Parliament dissolved (Government of Canada, 2023). In the meantime, Canada relies on existing laws (Privacy law, Human Rights Act, etc.) plus emerging guidance. Individual provinces also have specific guidance, such as Ontario's Consumer Reporting Act (Government of Ontario, 1990) and the passed-but-not-yet-enforced Consumer Protection Act (Government of Ontario, 2023).

Canada's banks and consumer regulators (the Office of the Superintendent of Financial Institutions (OSFI) and the Financial Consumer Agency of Canada) have actively identified AI risks. A joint risk report (Office of the Superintendent of Financial Institutions, 2024a) notes that Canadian banks are increasingly using AI for credit underwriting and adjudication, among other high-impact uses. OSFI has emphasized that AI/ML models amplify model risk and require robust management. The OSFI's model risk management guideline draft (E 23) explicitly warns that "risks [are] amplified by the surge in AI/ML analytics" and mandates strong life cycle controls (Office of the Superintendent of Financial Institutions, 2024b). These are principles-based expectations rather than new rules.

The proposed AI law (Artificial Intelligence and Data Act [AIDA]) is intended to target "high-impact" AI, and could impose obligations on bias mitigation and transparency. Like its EU counterpart, AIDA identifies credit risk as a high-impact activity. In parallel, updates to Canada's Consumer Privacy Protection Act will require consent for new data uses, indirectly affecting algorithmic scoring.

In practice, Canadian banks remain largely aligned with Basel and US norms: they must obey the Bank Act, the B-20 guidelines, and anti-discrimination laws when using AI in lending. The general trend is to integrate AI into existing model validation frameworks, an approach similar to international Basel guidelines, rather than to set prescriptive new AI-specific rules.

1.4.5 Singapore

The Monetary Authority of Singapore (MAS), the country's financial regulator, has taken a proactive, principles-based route to regulation that is unique to regulators worldwide. No new law specifically targets AI in credit, but the regulator has published detailed guidance and even a detailed Python package with functions to measure the fairness of models, which we present in Chapter 8.

More recently, in December 2024, the MAS published an information paper on the risk management of AI models (Monetary Authority of Singapore, 2024). The paper outlines good practices across three domains, governance, risk controls, and model development, strongly emphasizing fairness, transparency, and explainability. For example, MAS advises banks to establish cross-functional AI oversight committees, maintain detailed AI inventories, and include "fair, ethical, accountable, and transparent use of AI" in policy statements. During model development, the MAS recommends standards for data management and explicit checks for explainability and bias in AI models. It has also indicated that it will continue to work with firms to improve AI risk practices and may issue formal supervisory guidance in 2025.

In February 2024, Singapore's Personal Data Protection Commission (PDPC) published advisory guidelines on AI decision systems (Personal Data Protection Commission, 2024). These clarify that organizations must explain to consumers how personal data is used in automated decisions and obtain consent for data used to train AI. There is no absolute ban on alternative data, but companies must ensure compliance with relevant laws and manage discrimination risks.

In summary, MAS treats AI risk management similarly to other model risks (a tech-neutral stance), whereas PDPC ensures consumer rights in data use. Credit risk models using AI in Singapore must therefore meet the robust governance expectations of MAS and the transparency requirements of PDPA, but no new lending-specific statute has been enacted since mid-2025.

One complexity of modern AI use in banking is navigating the complex landscape at the intersection of banking regulation, privacy regulation, and AI-specific regulation. However, the line that most regulators and territories are following is to allow the use of complex models, as long as clear reasons can

be provided regarding the model performance and how it avoids unfair discrimination between borrowers. These restrictions are much stricter for consumers than in other segments and are universally focused on consumer credit scoring, not parameters, Internal Ratings-Based (IRB) models, or other uses. This leaves plenty of space for the use of deep learning and advanced AI in these segments. In what follows, we begin this journey with an introduction to neural networks.

1.5 TECHNICAL BACKGROUND

The history of AI is a fascinating chronicle of discoveries, inventions, and technological advances that have transformed how we interact with machines and how they integrate into our daily lives. From the dawn of computing to the present day, AI has evolved through different waves, each marked by significant milestones that have propelled the field to new horizons.

In 1943, Warren McCulloch, a neuroscientist, and Walter Pitts, a mathematical logician, published a pioneering paper titled "A Logical Calculus of the Ideas Immanent in Nervous Activity" (McCulloch and Pitts, 1943). This article is fundamental in the history of AI, as it was the first to conceptualize how neurons in the brain could perform complex calculations through a simple network based on Boolean logic and binary arithmetic. The main concept was that networks of simple neurons could theoretically compute any computable function, linking biology and computation.

The Dartmouth Conference of 1956 is considered the official birth of AI as a distinct field of study and research. It was at this event that the term "artificial intelligence" was first used, during a summer project organized by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon, who are considered some of the founding fathers of AI. The purpose of the conference was to gather experts from different disciplines to discuss the possibility of machines simulating various aspects of human intelligence (Muthukrishnan et al., 2020).

1.5.1 Early Days: The Perceptron

In 1957, psychologist Frank Rosenblatt invented the Perceptron (Rosenblatt, 1962), one of the first types of artificial neural networks. Influenced by the desire to understand how the human brain processes information and learns from experience, Rosenblatt designed the Perceptron as a simplified computational model of a biological neuron, capable of binary classification tasks. The Perceptron model consists of a single layer of artificial neurons that receive

inputs, weighs them with a series of weights, sums these weighted inputs, and then passes the result through a function to produce an output.

The first versions of machine learning (ML) models were designed for tabular data, which consist of a table or matrix, as in the case of credit scoring. This application considers each model input as a row or vector of information: the customer is described by a series of variables such as age and income. In supervised learning tasks, the process is guided by a target variable Y. Credit scoring is a classification task, as the target variable can take two values (binary classification): good payer (GP) or bad payer (BP). As shown in Figure 1.1, the table of explanatory variables or covariates is denoted by the letter **X.**

Formally, for N training samples of the form $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$, where $\mathbf{x}_i \in \mathbb{R}^p$ is a feature vector representing the i-th sample. For a binary classification problem, $y_i \in \{-1, +1\}$ is the class label of $\mathbf{x}_i$. A linear model such as the Perceptron constructs a hyperplane of the form $\mathbf{w}^\top \mathbf{x} + b = 0$, with $\mathbf{w} \in \mathbb{R}^p \backslash \{\mathbf{0}\}$ representing the *weights* of the model and $b \in \mathbb{R}$ the *bias*. The latter term captures the invariant part of the predictions, similar to the offset term in the traditional regression models.

The input layer of the Perceptron consists of p nodes that "transmit" the information of the p features $\mathbf{x}$ to an output node, with connections or edges $\{\mathbf{w}, b\}$. The output $\hat{y}$ is obtained as follows:

$$\hat{y} = sign\left(\mathbf{w}^\top \mathbf{x} + b\right) = sign\left(\sum_j^p w_j x_j + b\right),$$

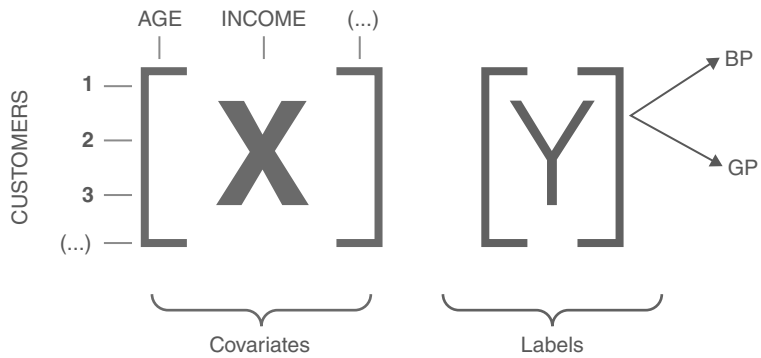

FIGURE 1.1 Example of a tabular dataset for supervised learning inspired by the credit scoring task.

where the sign function maps the weighted sum to either –1 or 1, transforming the numerical output into a class. The weights and the bias are obtained in such a way that the resulting hyperplane separates the two classes as well as possible.

The architecture of the Perceptron can be represented as in Figure 1.2. A neuron typically encompasses two functions: the summation, represented with the symbol Σ in Figure 1.2, which computes the weighted sum, and the activation function Φ, which introduces nonlinearity into the network, allowing it to learn complex patterns. In the Perceptron, the default activation function is the sign of the weighted sum, shown with a step function in Figure 1.2.

The choice of the activation function depends on the predictive task at hand. In regression problems, for example, the output range is a number between $(-\infty, \infty)$, therefore, the usual choice is the linear activation function (identity function): $\Phi_{id}(x) = x$. The activation function can be considered a *hyperparameter*, that is, an input whose value is set before the learning process begins and controls the behavior of the learning algorithm. Unlike the model parameters (the weights and the bias), which are learned during training, hyperparameters are set by the practitioner and remain constant throughout the training process. The following activation functions are commonly used in neural networks:

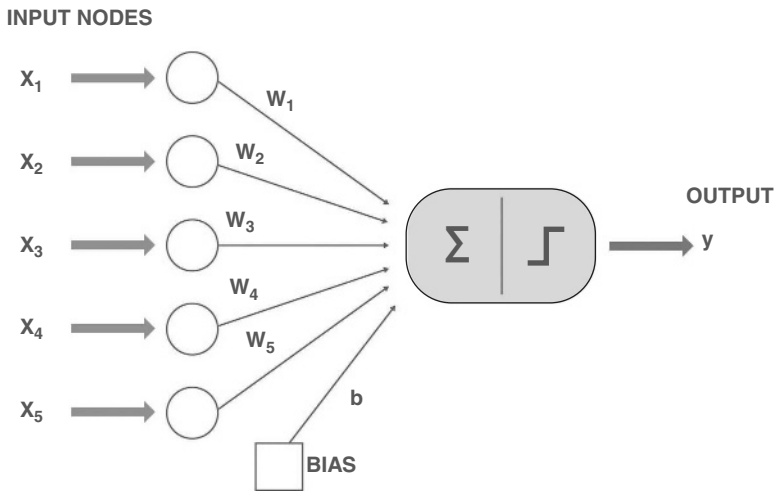

FIGURE 1.2 The architecture of the Perceptron model.

- **Sign Function:** With an output range of $\{-1, 1\}$, it is the basic approach for binary classification tasks.

$$\Phi_{sign}(x) = \begin{cases} -1 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$$

- **Sigmoid Function:** With an output range of $(0, 1)$, it provides a probabilistic outcome in binary classification tasks, similar to the logit function in the logistic regression model.

$$\Phi_{sigm}(x) = \frac{1}{1 + e^{-x}}$$

- **Softmax Function:** It provides a probabilistic outcome in multiclass classification tasks, having an output range of $(0, 1)$, summing to 1 across all classes.

$$\Phi_{softmax}(x) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}}$$

- **Rectified Linear Unit (ReLU):** With an output range of $[0, \infty)$, it provides a simple activation function with appealing properties for deep learning. ReLU has replaced the sigmoid function in modern machine learning:

$$\text{ReLU}(x) = \max(0, x)$$

The various activation functions described in this section are presented in Figure 1.3.

Along with the Perceptron, symbolic logic became popular in the early days of AI. It involves representing knowledge by combining abstract symbols and manipulating them with strict rules. McCarthy and his collaborators believed it was possible to create machines that could reason and solve problems if enough symbols and rules were combined (Muthukrishnan et al., 2020). However, the complexity of reality proved too much for strict rules, highlighting a significant limitation of this approach.

The focus on symbolic logic and problem-solving, along with the development of early AI programming languages like Lisp and the rise of expert systems (computer programs designed to emulate human decision-making in specific fields), are defining features of the "first wave" of AI. This era was marked by considerable optimism about AI's capabilities to mimic and eventually surpass human intelligence through rules and formal logic.

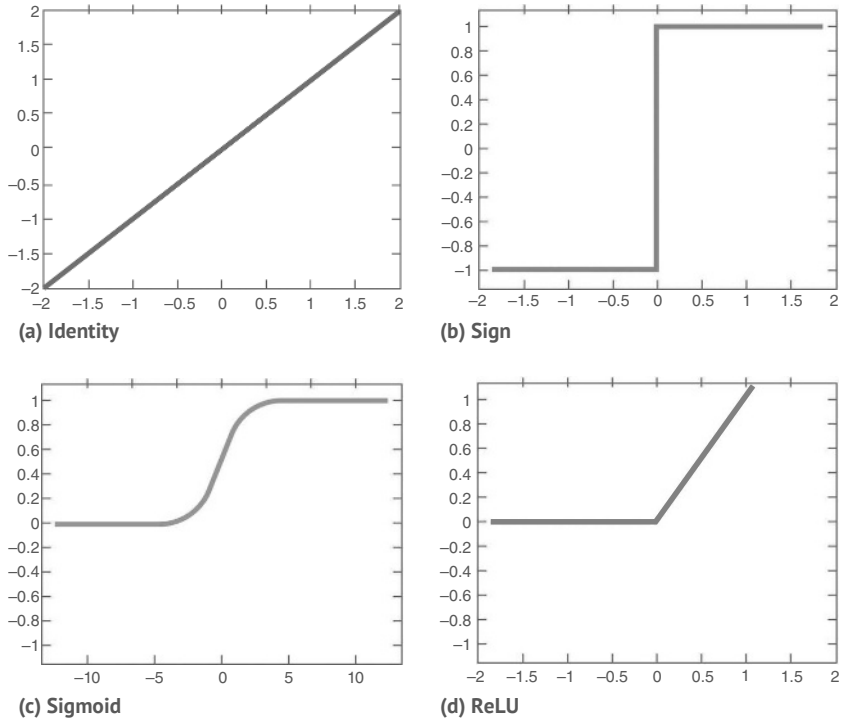

FIGURE 1.3 Graphical representation of various activation functions.

However, initial enthusiasm for the Perceptron was tempered by theoretical limitations exposed in 1969 by Marvin Minsky and Seymour Papert in their book *Perceptrons*. They showed that simple Perceptrons were incapable of solving linearly inseparable problems, such as the XOR function, significantly hindering neural network research and machine learning during the "First AI Winter" (Muthukrishnan et al., 2020).

1.5.2 The Multilayer Perceptron

Artificial neural networks saw a resurgence thanks to the Multilayer Perceptron (MLP) model and the introduction of the backpropagation algorithm in 1986 by Geoffrey Hinton, David E. Rumelhart, and Ronald J. Williams.[3] This algorithm became a fundamental method for training multilayer neural networks, allowing them to learn from errors and effectively adjust their internal

[3]Paul Werbos is often credited with the invention of backpropagation in his 1974 PhD thesis. However, the algorithm was significantly refined and popularized by Rumelhart et al. (1986).

weights. The introduction of backpropagation marked a revival in interest in neural networks, opening new possibilities for applications in pattern recognition, natural language processing, and computer vision. Geoffrey Hinton continued to be a key figure in AI and deep learning, contributing to innovations such as convolutional neural networks and Boltzmann machines. Hinton, along with Yoshua Bengio and Yann LeCun, was awarded the Turing Award in 2018 for his contributions to deep learning.

However, the greatest recognition for Hinton came in 2024 with the Nobel Prize in Physics, awarded mainly for the design of Boltzmann machines and the physical principles that inspired this model. He received this prize alongside John Hopfield, the creator of the Hopfield network, which is another physics-inspired model that laid the foundations for recurrent neural networks and helped revitalize the field of AI in the 1980s.

The main difference of MLP in relation to the Perceptron is the introduction of one or more "hidden layers" of neurons between the input layer (data) and the output layer (prediction), allowing it to model complex nonlinear functions with low predictive error. Its architecture consists of multiple layers of neurons arranged in a sequential manner, where each layer is fully connected to the next one. This intermediate layers are called "hidden" because the calculations are invisible for the modeler. In contrast, all the computations made in the Perceptron are done in the output layer, and are therefore visible for the modeler. The MLP architecture is presented in Figure 1.4.

The number of hidden layers and neurons in them are additional hyperparameters of the model, conferring flexibility to the approach. However, tuning multiple hyperparameters is a challenging and often time-consuming

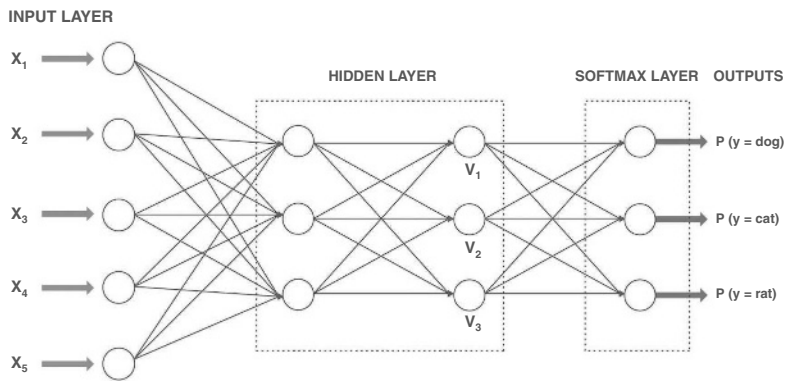

FIGURE 1.4 The architecture of the Multilayer Perceptron (MLP) model.

process that significantly impacts the model performance. Effective hyperparameter tuning can lead to models that generalize better on unseen data, while poor tuning can result in suboptimal performance.

More layers and neurons can model more complex functions but also increase the risk of *overfitting*. Overfitting is a modeling error that occurs when a model learns the details and noise in the training data to such an extent that it negatively impacts the model performance on new, unseen data. In other words, an overfitted model captures not only the underlying patterns but also the random fluctuations and noise present in the training data. This results in a model that performs well on the training data but poorly on validation or test data.

Theoretically, MLPs with a single hidden layer can approximate any continuous function given sufficient neurons and appropriate parameters (weights and biases). This is called the *Universal Approximation Theorem* and is a fundamental result in the theory of neural networks. This theorem highlights the expressive power of neural networks and provides a theoretical foundation for their ability to learn a wide variety of functions. Unfortunately, the theorem does not provide guidance on the number of neurons required. In practice, very large numbers of neurons might be needed, making deep networks (with multiple hidden layers) more practical for complex tasks.

Along with MLP, expert systems, which could answer questions or solve problems in a specific domain using logical rules derived from expert knowledge, became a major focus during the 1980s. These systems were limited to narrow domains of knowledge but proved useful and successful, as exemplified by eXpert CONfigurer (XCON), a system expert developed at the Carnegie Mellon University (CMU) for the Digital Equipment Corporation. This success spurred widespread corporate investment in AI.

Sutton (1988) laid the foundation for reinforcement learning by introducing a method where an agent learns to make decisions through interaction with an environment, receiving rewards or penalties, and aiming to maximize cumulative rewards. This framework, inspired by behavioral psychology, has since become a cornerstone of modern AI.

Despite these advancements, the "Second AI Winter" emerged in the late 1980s due to inflated expectations and the inability of expert systems to generalize beyond specific knowledge domains. This period saw reduced funding and support for AI projects and increased skepticism about AI's ability to emulate human intelligence.

From the 1990s onward, other forms of machine learning, such as Support Vector Machines (SVMs) and ensemble methods like random forests, began to outperform artificial neural networks (ANNs) in many tasks. These techniques,

based on statistical learning theory and other principles, helped revitalize the field and led to significant practical applications.

1.5.3 Training MLP: Backpropagation

The optimization of weights in a multilayer network was made feasible by the backpropagation algorithm. This algorithm efficiently calculates gradients using the chain rule of differential calculus, which are then used to update the weights through gradient descent. This process allows the neural network to "learn" from the errors in its outputs by adjusting its weights to minimize the error between the predicted outputs and the actual targets.

The backpropagation algorithm is divided in two stages: The *feedforward* step passes the input variables through the network layers to produce an output, while the *backpropagation* step updates the network parameters to minimize the training error.

For the feedforward step, the input data is first fed into the network through the input layer. Next, the input is then propagated forward through each hidden layer. For each neuron m in hidden layer l:

$$a_m^{(l)} = \Phi\left(\sum_n w_{mn}^{(l)} a_n^{(l-1)} + b_m^{(l)}\right), \tag{1.1}$$

where $w_{mn}^{(l)}$ and $b_m^{(l)}$ are the model parameters, and $a_n^{(l-1)}$ represent the activations from the previous layer. This process continues through all hidden layers until the final output layer is reached. The neurons in the output layer produce the final predictions of the network $\hat{y}$, using a similar process of weighted sums and activation functions.

After calculating the functional form of the network in a forward manner, the error at the output layer is computed using a *loss function*. The loss function $\mathcal{L}$ depends on the predictive task:

- **Regression tasks:** The most common choice is the mean squared error (MSE).

$$\mathcal{L}_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

- **Binary classification tasks:** The binary cross-entropy (CE) loss is usually considered.

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right]$$

- **Multiclass classification tasks:** The general cross-entropy loss is usually considered. Note that this function is equivalent to the binary CE when two classes are considered ($K = 2$).

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik}\log(\hat{y}_{ik}).$$

Then, given a suitable loss function $\mathcal{L}$, the error for each output neuron k can be represented as:

$$\delta_k^{(L)} = \frac{\partial \mathcal{L}}{\partial z_k^{(L)}},$$

where L denotes the output layer. The error is then propagated backward through the network, layer by layer (backpropagation step). For each neuron m in hidden layer l, the error $\delta_m^{(l)}$ is computed using the error from the subsequent layer $(l+1)$:

$$\delta_m^{(l)} = \left(\sum_n \delta_n^{(l+1)} w_{mn}^{(l+1)}\right)\Phi'(z_m^{(l)})$$

where $\Phi'(\cdot)$ is the derivative of the activation function, and $z_m^{(l)}$ is the argument of this function (the weighted sum of activations from the previous layer plus the neuron's bias).

Next, the gradients of the loss function with respect to the weights and biases are calculated. They indicate the direction in which the loss function is increasing the most rapidly. Moving in the opposite direction of the gradient decreases the value of the loss function. The following expressions define the gradients for the weights $w_{mn}^{(l)}$ and biases $b_m^{(l)}$:

$$\frac{\partial \mathcal{L}}{\partial w_{mn}^{(l)}} = a_n^{(l-1)}\delta_m^{(l)},$$

$$\frac{\partial \mathcal{L}}{\partial b_m^{(l)}} = \delta_m^{(l)}.$$

Finally, using gradient descent (or one of its variants), the weights and biases are updated to minimize the loss function. The updates are typically performed as:

$$w_{mn}^{(l)} \leftarrow w_{mn}^{(l)} - \eta\frac{\partial \mathcal{L}}{\partial w_{mn}^{(l)}}$$

$$b_m^{(l)} \leftarrow b_m^{(l)} - \eta \frac{\partial \mathcal{L}}{\partial b_m^{(l)}}$$

where η is the *learning rate*, which is a hyperparameter that controls the step size at each iteration while moving toward a minimum of the loss function. An additional hyperparameter is *momentum*, which is used to accelerate the training process and improve convergence in the backpropagation algorithm. The idea is to give the optimizer inertia with a decay factor, allowing it to maintain its direction even in the presence of small gradients or noise.

To simplify the notation, consider a vector θ that concatenates all the weights and biases from all layers of the neural network. The update rule for the model parameters with momentum becomes:

$$\begin{aligned} v_t &= \beta v_{t-1} + (1-\beta)\nabla_{\theta_t}\mathcal{L} \\ \theta_{t+1} &= \theta_t - \eta v_t, \end{aligned}$$

where v_t is the momentum vector at iteration t, ∇_{θ_t} is the gradient (the vector of partial derivatives) associated with the parameter vector θ, and β is the momentum coefficient. In traditional MLP, the initial parameters θ_0 are defined at random. The previous equations can also be written in an alternative form:

$$v_t = \beta v_{t-1} + \nabla_{\theta_t}\mathcal{L} \tag{1.2}$$

$$\theta_{t+1} = \theta_t - \eta v_t. \tag{1.3}$$

Here, the gradient term in the momentum update equation directly includes the current gradient without the factor $1-\beta$, which is a common practice in MLP implementations.

In summary, the backpropagation algorithm encompasses two steps. In the forward pass, the model parameters remain fixed and the predicted values $\hat{y}$ are obtained using Equation (1.1). In the backward pass, errors $\delta_m^{(l)}$ are calculated and backpropagated, and then utilized to update the model parameters by Equation (1.2).

The MLP networks include several hyperparameters, making them challenging to tune. First, the number of hidden layers and hidden units (neurons) needs to be defined. The backpropagation algorithm includes the learning rate and the momentum as tuning hyperparameters for the update rule. In addition, the number of full iterations, or *epochs T*, must be defined. An epoch represents one complete pass through the entire training dataset. During one epoch, the model sees every example in the training set once and updates its parameters based on the loss function.

Finally, an additional hyperparameter is the regularization term or *weight decay*. MLP networks with many weights tend to overfit the data, learning both

the patterns and the intrinsic noise. Overfitted networks perform well on the training data but fail to generalize to new, unseen data. One of the first solutions to this issue was *early stopping*, which involves stopping the training process after a small number of epochs T to prevent the model from reaching the global minimum. A more formal approach is to introduce a regularization term that shrinks the parameters toward zero. Instead of minimizing a single objective given by the loss function $\mathcal{L}$, a regularized MLP network balances two objectives: model fit and shrinkage.

Two common shrinkage strategies are L2 regularization (Ridge) and L1 regularization (LASSO). The first minimizes the squared Euclidean norm of the parameters:

$$J(\theta) = \mathcal{L}(\theta) + \frac{\lambda}{2} \|\theta\|_2^2,$$

where $\lambda > 0$ is a tuning hyperparameter. The LASSO, on the other hand, minimizes the 1-norm of the parameters:

$$J(\theta) = \mathcal{L}(\theta) + \lambda \|\theta\|_1 .$$

1.5.4 From Multilayer to Deep Learning

The origins of deep learning trace back to the late 1980s with the development of convolutional neural networks (CNNs) by Yann LeCun and others for automatic image analysis. However, this model required an enormous amount of computational resources and images to function effectively.

The revolution of deep learning mainly enables the direct modeling of sophisticated data, bypassing the need for extensive feature engineering, which often results in information loss. By sophisticated data, we refer to data that go beyond the traditional tabular structure. When analyzing images, an additional dimension to the tabular setting must be added, as each input is a matrix or table, resulting in a "cube" of information. The mathematical concept for this cube is the *tensor*, as represented in Figure 1.5. Note that a tensor can have any number of dimensions, though 0D (scalar), 1D (vector), and 2D (matrix) tensors are so commonly used that they are given special names. Tensors can also have more than three dimensions and may take more complex forms, such as those used in natural language processing or other structured data types.

The 2010s saw a significant AI boom, largely driven by the increase in computational power through more powerful GPUs and other technical advancements in terms of hardware. This computational power enabled the training of larger and more complex neural networks, a crucial factor behind many recent AI advancements. Furthermore, massive digitization and the generation

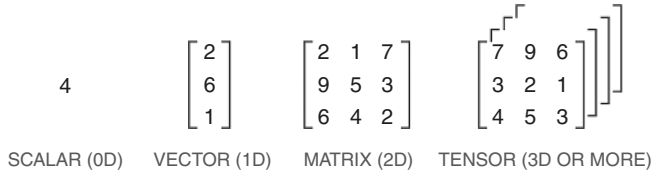

FIGURE 1.5 Illustration of different data representations: scalar (0D), vector (1D), matrix (2D), and a three-dimensional tensor. While tensors can have any number of dimensions, the 0D, 1D, and 2D cases are so commonly used that they are given special names.

of large volumes of data have provided the "fuel" necessary to train deep learning models.

Apart from advances in computational power, the last few decades have seen significant technical advancements in the optimization processes behind artificial neural networks, enhancing the efficiency and effectiveness of the backpropagation algorithm. Among these developments, which will be detailed in the next chapters, are the following:

- The introduction of the ReLU activation function, which is more effective and efficient than the sigmoid function. ReLU helps address stability issues related to gradient computation in deep networks, particularly the *vanishing gradient* and *exploding gradient* problems.
- The development of *stochastic gradient descent* (SGD) as a faster alternative to computing the exact gradient of the loss function over the entire dataset. SGD updates the model parameters using only a small batch of samples at each iteration, making the updates more frequent and computationally feasible for large datasets.
- Enhanced SGD methods, such as Adaptive Moment Estimation (ADAM), have been proposed to facilitate the search for optimal hyperparameters. Adam (Kingma and Ba, 2014), in particular, adaptively finds the learning rate and momentum, improving the optimization process.
- The key methodological advance of *transfer learning*, which involves reusing a pre-trained model developed for one task as the starting point for a model on a second task. This approach is especially beneficial in fields like text analytics and computer vision, where training models from scratch typically requires large amounts of data and significant computational resources.

There has also been a substantial increase in investment and funding from both public and private sources for AI research and development. A notable

example is DeepMind Technologies, founded in 2010, which gained worldwide fame in March 2016 when its model AlphaGo, a computer program that plays the board game Go, defeated Lee Sedol, one of the world's top Go players. This was significant because Go is a particularly complex game with a large number of possible moves, making it resistant to previous AI and brute-force approaches. AlphaGo used deep neural networks and reinforcement learning algorithms to learn game strategies from human games and then improved by playing against itself.

In January 2014, Google acquired DeepMind for a reported sum of more than $500 million, reflecting Google's growing interest in AI and recognizing DeepMind's potential to make significant contributions to the field.

DeepMind's contribution has been much deeper than just AlphaGo and other methodological advances. In 2020, Demis Hassabis and John Jumper presented an AI model called AlphaFold2, which allows the prediction of the 3D structure of proteins based on amino acid sequences. With the help of this model, scientists have managed to predict the structure of nearly all the proteins identified to date. Among its many scientific applications, researchers can now better understand antibiotic resistance and generate images of enzymes capable of breaking down plastics. This work led them to win the 2024 Nobel Prize in Chemistry, along with David Baker, a biochemist and pioneer in protein design based on amino acid sequences. In 2003, Baker succeeded in designing a new protein unlike any existing one, creating a new field of research focused on designing proteins for novel vaccines and medications.

In 2014, Ian Goodfellow and his colleagues introduced the concept of Generative Adversarial Networks (GANs) in a seminal paper, marking a milestone in deep learning and AI. GANs are an innovative approach to generative learning that enables the creation of highly realistic images, texts, music, and other data types that were previously difficult or impossible to generate using traditional methods. This work laid the foundation for Generative AI (GenAI).

A concrete and significant milestone in AI applied to health care is the development and approval of IDx-DR in 2018, the first AI system authorized by the US Food and Drug Administration (US FDA) for the automatic detection of diabetic retinopathy, a complication of the eyes that can lead to blindness if not treated. IDx-DR analyzes retinal images captured with specific retinal cameras and uses AI algorithms to identify signs of diabetic retinopathy.

From 2022 to 2023, Generative AI experienced significant advances with the launch and popularization of technologies such as MidJourney, Stable Diffusion, and DALL-E for image generation, and OpenAI's ChatGPT for text

generation. This marks an era of democratized access to powerful and versatile AI tools. These innovations have allowed users, regardless of their technical expertise, to explore and create complex multimedia content, such as text, images, audio, and video. The major technical advancements that made this boom possible were the development of attention mechanisms and deep learning models based on transformers (Vaswani et al., 2017).

This introduction provided the background necessary to understand the rest of the book. We are ready now to continue with our first data type, and first family of models: images and convolutional neural networks (CNNs).