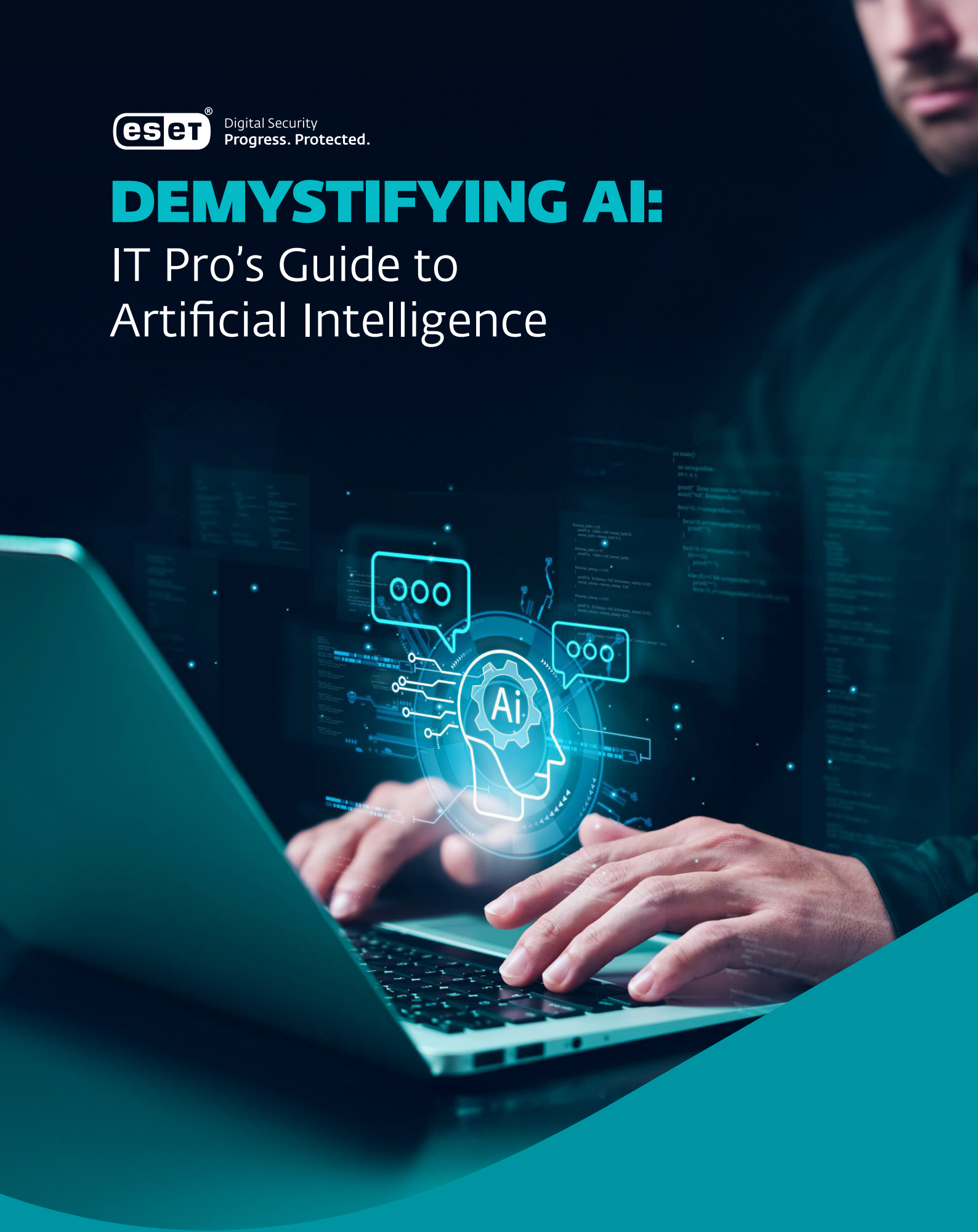
Digital Security
Progress. Protected.

DEMYSTIFYING AI:

IT Pro's Guide to Artificial Intelligence



INTRODUCTION

Artificial intelligence, or AI, has been one of the hottest topics in the security industry over the past four or five years, and over the past year, the buzz has amplified. The public availability of generative AI, most of all ChatGPT, brought AI out of the labs, democratized it, and captured imaginations everywhere with its almost-human ability to understand prompts and generate easily readable, comprehensible text.

Among security product leaders:[1]

- **More than 50% are using supervised machine learning already**
- **More than 80% are actively developing or have plans to integrate generative AI, specifically large-language models**

Security software providers, however, have recognized there is a dark side to generative AI: the potential for threat actors to use the technology to generate new malware code and more-convincing phishing emails at higher speed and scale. Indeed, there was a very sharp uptick in mentions of generative AI when ChatGPT was launched, and one threat actor claimed to be able to use generative AI to create malware from research publications.[2] In response, security product leaders are accelerating their adoption and implementation of AI technologies to counter an increased volume and sophistication of threats. AI technologies can help security analysts by doing two things that AI does really well: (1) sift through and make sense of reams of data faster than a human can, at a scale that's beyond human capability; and (2) automate parts of the job to free up analysts for higher-value work.

As more AI tools are being added to the security toolbox, we put this guide together to crack open the "magic black box" mentality that currently surrounds AI and put it into perspective. Generative AI seems shiny and new, but other forms of AI are being used in all sorts of ways today. We take them for granted, and we don't think of those applications as AI. ChatGPT and its kin are just the latest development in a long continuum of advances that stretch back decades. (For instance, ESET products have been using two different forms of advanced machine learning for years — a high-powered detection engine in the cloud and a lightweight version on the endpoint.)

MACHINE LEARNING

Technically, AI is a subset of the broader field of data science, and machine learning is a subset of AI. The term "machine learning" has been around since 1959, and it involves training a model on a set of data so it can learn how to predict future events.

Not every form of AI involves machine learning, but many of the more significant, successful and widely used applications of AI draw from the technique. In the popular view, the technical distinction between the two is immaterial: AI and machine learning are one and the same.

Machine learning practitioners recognize four broad types of learning models.

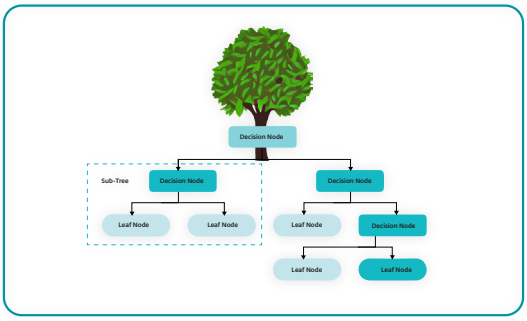
1. In supervised learning, the data has been labeled by humans. Data scientists have looked at the data set to determine the most relevant patterns and relationships.
2. In unsupervised learning, the model itself discovers the patterns in the data.
3. In semi-supervised learning, a small amount of labeled data and a large amount of unlabeled data are used to train a predictive model.
4. Reinforcement learning tunes the model after deployment, using real-world data to improve the model's predictive power.

1 Gartner, Emerging Tech: Top Use Cases for AI in Threat Detection, Investigation and Response, Travis Lee, Matt Milone, Elizabeth Kim, John Collins, October 27, 2023
2 Bain and Company, Technology Report 2023

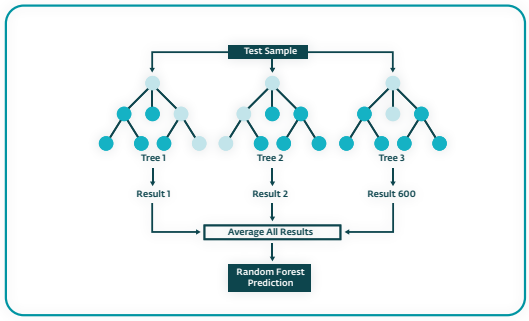
AI ALGORITHMS

The goal of machine learning is to arrive at a function that provides the most precise prediction (one or more output values) based on a set of input values. In security, the output could be identifying that a sample is malicious or that a security event isn't isolated but is part of an incident. If there's an art to AI, it's in knowing which choice of a machine learning algorithm — or combination of algorithms — is the best choice for accomplishing the task.

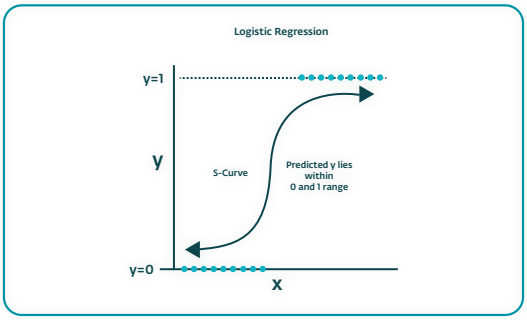
Decision trees are one of the oldest, simplest machine learning models: The input value is tested against a series of binary yes-no decisions to arrive at the output.



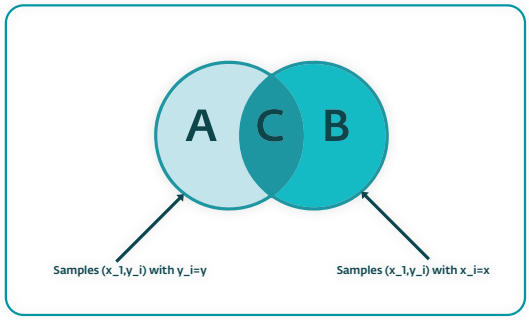
Random forest builds multiple decision trees based on different subsets of the data and takes the aggregate of their outputs to arrive at the final decision.



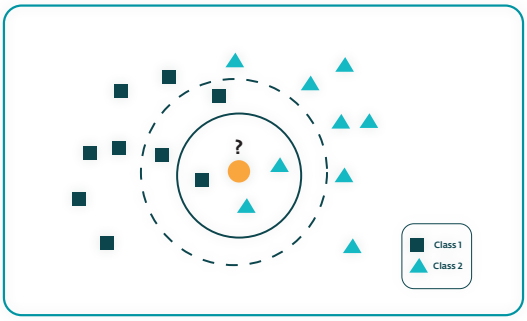
Logistic regression is commonly used for classification when the output is a binary yes-no result, and it identifies the independent variables, and their relative weights, that will best predict the output.



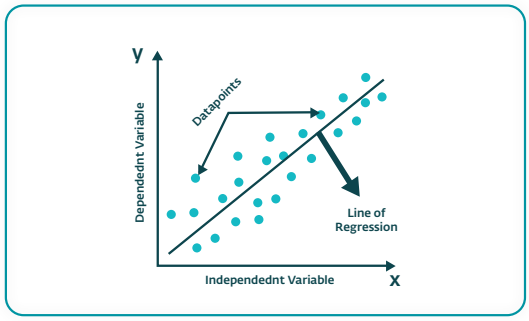
Naïve Bayes can be used for both binary and multiple-class classifications and is based on Bayes' theorem, which describes the probability of an event based on conditions that might be related to it.



K-nearest neighbors works on the assumption that similar data items exist near each other and predicts the outcome value by checking the entire data set for data nodes with similar values.

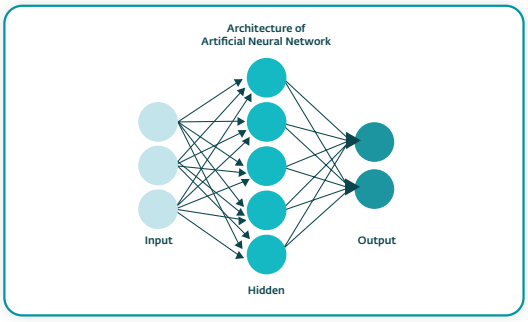


Linear regression aims to arrive at a linear equation that best predicts an outcome (dependent variable) based on the input (independent variable) by adjusting the number and weight of the coefficients in the equation.



ARTIFICIAL NEURAL NETWORKS

Artificial neural networks (ANNs) take their inspiration from and are an attempt to emulate the way the human brain works. The term "artificial" simply makes it clear that the neural network we're talking about here is no match for its biological, human counterpart. Nobody has figured out — yet — how to come up with a structure and an algorithm that works exactly like the human brain works, with its densely packed, massively interconnected neurons.



While most of the AI algorithms on the previous page are supervised models, ANNs are capable of supervised, unsupervised and semi-supervised learning. They were invented in the 1950s, but the recent availability of massive data sets and cloud-borne, highly scalable computing power for training them is the most significant factor in their widespread use.

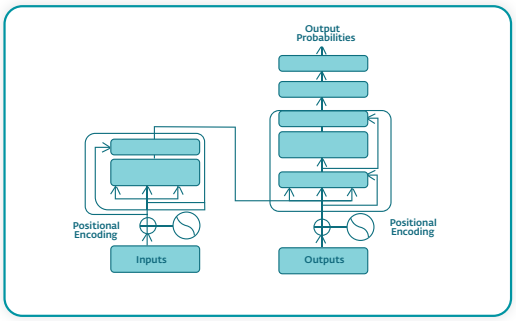
An ANN has an input layer, an output layer, and one or more "hidden layers" where the processing happens. Each hidden layer is responsible for handling one step in the processing. The nodes in each layer accept values from the previous layer, assign weights to them and pass them on to the nodes in the next layer. A weight of more than one indicates that the layer's result is more significant in the final results, while a weight less than one indicates less significance. The processing sequence is not simply forward only; "backpropagation," developed in the 1980s, allows an ANN with a large number of hidden layers to be trained. This is a highly simplified description, and a more complete explanation could easily take volumes, but it gives you a picture of how an ANN learns which combination of attributes and weights most accurately predicts an outcome.

ESET currently uses a more advanced form of ANN, called a recurrent neural network, and a further improvement called "long short-term memory" that allows the model to retain and "remember" key information as it flows through the network. ESET also uses a "deep neural network" and "deep learning," which refer to an ANN with a large number of hidden layers and the training of a model that uses such a network, respectively.

GENERATIVE AI

In 2017, a group of Google researchers published a paper that laid the groundwork for what we know today as generative AI.[3] They described an architecture they termed a "transformer" that applies AI attention mechanisms to draw global dependencies between input and output. Relying solely on attention rather than other mechanisms effectively solves for a problem encountered by previous neural networks: how to deal with longer strings of text. Attention:

- Enables encoding the meaning and structure of language directly, without requiring an additional neural network
- Processes each word or token in parallel for efficiency and scalability
- Adds context by emphasizing the weight of related words
- Allows the model to capture subtle details about meaning and interaction



Source: Vaswani A. et. al, Attention Is All You Need, 2017.

1 Vaswani A. et. al, Attention Is All You Need, 2017.

The implications went far beyond language processing; transformer architecture unlocked a way to model virtually any type of information. Generative AI — AI that can output content based on prompts supplied by humans — is currently available in models that generate images, speech, music and code as well as text where they are termed large language models, or LLMs.

Development of an LLM involves pre-training the model in an unsupervised manner on a large corpus of data to learn the syntax and semantics of natural language. (In ChatGPT's case, this included 45 terabytes of text data gathered from the internet, supplemented with several databases that contain human dialogue to support a conversational style.)

Then deep learning using the transformer architecture enables the LLM to understand and recognize the relationships and connections between words and concepts. Once the training is complete, the model is ready to generate responses to human prompts.

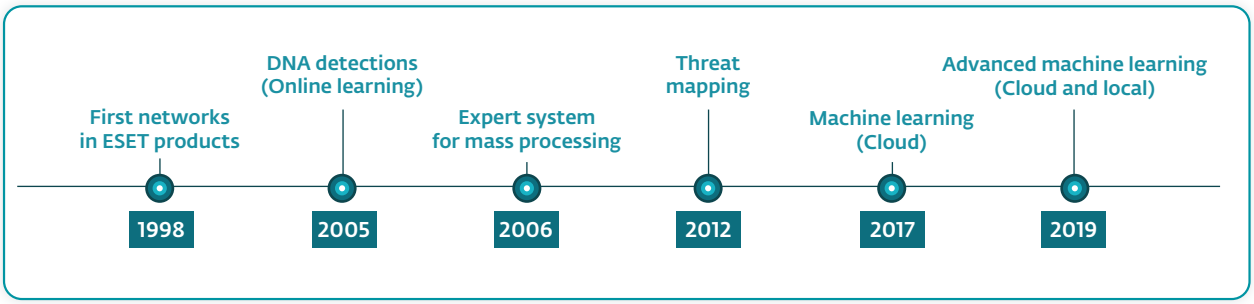
Now you know why, in the name ChatGPT, GPT stands for "generative pre-trained transformer."

AI IN ESET PRODUCTS

In the mid-1990s, ESET engineers recognized the potential for neural networks to detect rapidly evolving threats with greater speed. Since 1998, when these new detection algorithms were first introduced into ESET products, our AI journey has continued through a series of advances. Some of them are found directly in our products, while others are in our research labs where they are used to extract new insights and unlock new methods for applying the intelligence that can be gleaned from our massive threat database.

Today, AI in ESET products is used to:

- Detect and block threats in real time
- Combat new threats and ensure continuous threat intelligence updates via our global ESET LiveGrid reputation and response system
- Act on behavioral analysis and classify samples as clean, potentially unwanted or malicious using either a lightweight machine learning engine on the endpoint or a more robust algorithm in the cloud
- Power the detection-correlation engine in ESET Inspect, our XDR-enabling solution that ingests log data and automatically generates security incidents for attention and action by security analysts



4 Gartner, Emerging Tech – Top Use Cases for AI in Threat Detection, Investigation and Response, Travis Lee and Team, 2023, ID: G0079125

A PARTING WORD

As we indicated earlier, this was a quick introduction to AI — fully explaining the under-the-hood details and variations of the different algorithms would take volumes, and reams of academic papers cover all the variations. We hope this guide provided a better understanding of how they work and removed some of the mystery.

If there's one thing we would like you to take away, it's this: Generative AI seems shiny and new, but it's just the latest development built on a vast body of previous work. According to Gartner: "For years, security vendors have been integrating AI into their products and threat detection processes, to enable them to identify anomalies and patterns that surpass human capabilities at scale. In fact, over 50% of interviewed security service and tech providers claim to already be using supervised ML-based AI coupled with supervised training to enhance their threat detection abilities."[4]

AI algorithms have been in use for years, handling loan approvals, detecting payment card fraud, guiding retail decisions about shelf placement, suggesting purchase decisions based on past purchase history and hundreds of other examples. People don't usually think of these examples as AI, but they are because they emulate human judgment in an automated way.

It's likely that some years down the line, we won't think of generative AI as AI either. It's going to be an accepted part of the way we interact with technology and part of everyday life. For security analysts, in particular, it will automate some of the more repetitive and tedious parts of the job and bring sharper focus to the events and incidents that matter.

About ESET

AI-NATIVE PREVENTION FOR TOMORROW'S THREATS

Stay one step ahead of known and emerging cyber threats with our **AI-native, prevention-first approach**. ESET combines the power of AI and human expertise to make protection easy and effective.

Developed over 30 years, ESET's best-in-class protection is powered by our **in-house global cyber threat intelligence**, including our **extensive R&D network** led by industry-acclaimed researchers.

ESET PROTECT, our scalable, cloud-first XDR cybersecurity platform combines next-gen prevention, detection and proactive threat-hunting capabilities with a wide range of security services, including **managed detection and response** (MDR). Our highly customizable, integration-ready solutions support all deployment methods, include local support and have minimal impact on performance. They identify and neutralize known and emerging threats before they can be executed, support business continuity, and reduce the cost of implementation and management.

Our mission isn't just to stop attacks in their tracks; it's to prevent them from ever happening. ESET protects your business so you can unlock the full potential of technology.

ESET® Digital Security
Progress. Protected.