knowbe4

The KnowBe4 Approach to Human Risk Management

Meet AIDA: Artificial Intelligence Defense Agents

WWW.KNOWBE4.COM

KnowBe4, Inc. | 33 N Garden Ave, Suite 1200, Clearwater, FL 33755
Tel: 855-KNOWBE4 (566-9234) | Email: Sales@KnowBe4.com

CR-06120526

Introduction

The rise of AI-powered cyber attacks is making every infosec professional take notice.

More than 95% of cybersecurity professionals say AI-generated phishing is harder to detect, and Microsoft reports that AI-generated phishing emails are 4.5x more successful than manually created ones.

The problem isn't awareness of the threat. It's the reality that traditional security awareness programs can't adapt fast enough to keep pace.

This technological advancement in the hands of bad actors has created a new breed of highly convincing social engineering attacks that generic, one-size-fits-all security awareness training (SAT) struggles to combat.

It's time to fight AI with AI. Meet KnowBe4 AIDA — Artificial Intelligence Defense Agents. AIDA is a suite of AI-powered agents that up-levels your approach to human risk management (HRM) by leveraging multiple AI technologies to create personalized, adaptive and highly effective user training that actually changes behavior.

In this whitepaper, we'll discuss how each of the AIDA agents can help improve your HRM strategy and explore use cases.

Inside this Report

Let's Talk About AIDA

AIDA is a suite of always-on agents that up-levels your approach to HRM by leveraging multiple AI technologies to create personalized, adaptive and highly effective training for all of your users that actually changes behavior.

By automating template generation, training, phishing simulations and reporting, AIDA reduces the administrative burden on your security teams so they can focus on protecting your network.

One Agent to Rule Them All

With AIDA Orchestration, this automation extends to full program management—continuously assessing individual user risk and automatically delivering the right tests and training at the right time. This AI-native suite of agents empowers your users to turn into active threat detectors, strengthening your organization's defense.

AIDA continuously learns and adapts to your organization's unique needs, ensuring maximum impact and ROI for your security awareness efforts. One example: by reducing the time it takes your admins to create training content from an average of seven days to mere seconds, AIDA dramatically frees your security team's time.

Here's a summary of the agents that comprise the AIDA suite (with more detail and use cases later in this whitepaper):

Orchestration Agent

AIDA automates program administration with an "always-on" approach, continuously assessing user risk to automatically create and schedule individualized phishing security tests, remedial training and ongoing training campaigns.

- **Remedial Training Agent:** Automatically assigns the most relevant and engaging content when your users fail a simulated phishing test.
- **Ongoing Training Agent:** Continuously addresses knowledge gaps across your organization and automatically assigns personalized training content to each learner.
- **Phishing Agent:** Eliminates the manual burden of security testing by autonomously creating, customizing and delivering sophisticated phishing simulations tailored to each of your users' roles, risk profiles and previous interactions.

Template Generation Agent

Leveraging generative AI, AIDA creates highly realistic phishing templates that can mirror current attack vectors. Social Engineering Indicators, or red flags, are based upon the NIST Phish Scale Framework.

Callback Template Generation Agent:

Uses AI to create phishing templates that allow you to test your users on how likely they are to fall for callback phishing scams, which use a combination of phishing and vishing techniques.

Recommended Landing Pages Agent:

Automatically suggests contextually appropriate landing pages to accompany your AIDA-generated phishing templates, ensuring a complete and educational testing experience that reinforces learning objectives.

Knowledge Refresher Agent

AIDA delivers bite-sized knowledge refreshers at optimal intervals, ensuring your users actually apply critical security concepts.

Policy Quiz Agent

AIDA generates intelligent quizzes based on your organization's specific security and compliance policies. This ensures your users not only acknowledge, but truly understand the guidelines they're expected to follow.

Deepfake Training Content Agent

Generates custom deepfake training content featuring a leader from your own organization. This AI-generated content demonstrates how convincing AI-powered social engineering has become and delivers clear, actionable guidance on how to detect these attacks.

Custom SAPA Agent

The Custom Security Awareness Proficiency Assessment (SAPA) Agent creates custom risk assessments using your organization's specific documentation and policies, not generic awareness, to assess your users. Uses AI to select relevant questions from a large bank to build an organization-specific assessment based on your environment.

The Human Risk Management Connection

HRM is now the primary way to address the ongoing need for strong security cultures in organizations of all sizes. The concept focuses on more than just SAT delivered at regular intervals. The goal is a positive security culture through three primary tactics:

1. **Quantifying human risk through detecting and measuring human behaviors**
2. **Using these measurements to deploy relevant, tailored training content**
3. **Establishing recurring educational knowledge refreshers to ensure users know how to protect themselves against the most prevalent and impactful cyber risk**

Traditional SAT is no longer effective in your complex work environment. Once-per-year training events don't cut it. Plus, users' schedules are only getting busier, so the training that is delivered needs to make the most of their time.

Adopting an HRM strategy empowers you to deliver relevant, engaging content that changes user behavior and resonates across different roles, departments and even languages. AIDA is the backbone of this approach.

Continue reading for a deeper dive into the individual agents.

The Agents of AIDA

Before we talk about the agents, let's explore a key KnowBe4 feature that underpins them all: **KnowBe4's SmartRisk Agent™**. This capability provides actionable data and metrics, called the **Risk Score**, to help you better understand your organization's security strengths and weaknesses.

SmartRisk Agent leverages end user behavioral data from across KnowBe4's products to help measure cyber risk in humans. This agent's multidimensional Risk Scores are designed to make it easier for you to see potential problems at the user, group and organizational levels. You'll get more comprehensive insights the more KnowBe4 products you use.

Let's dive into each of the agent's capabilities and use cases.

Orchestration Agent

What It Does

AIDA Orchestration is an always-on, AI-driven system that automatically manages and personalizes KnowBe4 Phishing Security Tests (PSTs) and SAT to reduce user risk. AIDA uses AI to intelligently generate phishing templates and automatically sends those phishing tests to your users.

AIDA automatically scales the difficulty of each test based on user performance. The system uses personalized content and realistic logo changes to keep simulations relevant and appropriately challenging.

Even with this system enabled, you will retain full control to fine-tune the system's behavior and monitor all activity.

How To Use It

With the Orchestration agent, you define "Plans" to constrain testing and training frequency for user groups, while AIDA independently decides who to test, which attack vectors to use, and optimal timing to efficiently reduce organizational risk.

The Orchestration agent manages these three agents to build these plans:

- **Remedial Training Agent:** Ensures that every one of your users, regardless of position or location, receives training tailored to their specific risks, needs and learning style. This feature helps maximize retention and real-world application of your security best practices.
- **Ongoing Training Agent:** Automatically assign regular training to users based on their unique risk profiles.
- **Phishing Agent:** Delivers role and person-specific phishing exercises

Use Cases

→ **Tackling Phishing Vulnerabilities**
Take user responses or failure to specific simulated phishing tests and automatically assign training that teaches users about those pinpointed phishing strategies and tactics before they see them in the wild.

→ **Onboarding New Employees**
Assess new hire Risk Score and suggest SAT tailored to their existing knowledge base by using targeted Smart Groups.

Template Generation Agent

What It Does

The Template Generation Agent produces highly convincing phishing emails tailored to your organization's specific needs and risk profile using the power of baked-in generative AI. It can craft diverse, multi-lingual phishing scenarios with varying difficulty levels, tones and attack vectors, ensuring that your users are prepared for the full spectrum of potential threats they might encounter in the wild.

How To Use It

Adjustments including language, locale, tone and desired attack vector are all at your fingertips. Fine-tune simulated phishing email difficulty by selecting specific Social Engineering Indicators (SEIs) or choosing a predefined difficulty level. Once generated, templates can be previewed, edited and saved for future use, saving you valuable time in creating and managing a robust phishing simulation program.

Use Cases

- **Multi-lingual Phishing Tests**
  Quickly draft and deploy phishing emails in multiple languages to support a global workforce. The template generator can even adapt email text to a specific region with local expressions, nuances and cultural styles.

- **Rapid Response to Emerging Threats**
  Seeing a brand new, in-the-wild attack? A few minutes with the template generator is all you need to spin up a duplicate simulated phishing attack to prepare your workforce for what's to come.

Callback Template Generation Agent

What It Does

The Callback Template Generation Agent creates simulated callback phishing attacks where the primary goal is to prompt the user to call a specific phone number rather than click a link. The AIDA agent generates audio greetings, responses (heard after the user interacts with the greeting) and translations for both, alongside the email body text.

How To Use It

Callback templates are created just like email-based phishing templates. Choose from "Quick Create" for rapid generation or fine tune tone, difficulty, personalization and languages. Once generated, the greeting and response scripts are fully editable to maximize personalization.

Use Cases

- **Urgent Invoice Overdue**
  Send specific users an email claiming a payment is past due or a subscription is about to renew automatically. The goal would be to drive the user to call the provided number to dispute the charge or cancel the service. AIDA creates a high-pressure email along with a simulated customer service audio greeting for when the user calls.

- **CEO Fraud** (Executive Impersonation)
  The user receives a message appearing to come from the CEO or a high-level executive requests the user's immediate attention on a time-sensitive matter. The goal would be to convince the user to call a number to "verify" details or receive verbal instructions. AIDA drafts a tone-appropriate email (e.g., serious but appreciative) and generates a voice response that mimics an executive or their assistant handling the call.

Recommended Landing Pages Agent

What It Does

The Recommended Landing Pages Agent works in tandem with the Template Generation Agent to suggest relevant, believable web pages linked from simulated phishing emails. Together these agents perform template generation and optimization in seconds, not hours, supporting a cohesive learning experience for the user with less manual work for you.

How To Use It

When the Template Generation agent is used to create a phishing template with a link or QR Code attack vector, the Recommended Landing Page agent automatically selects a landing page for that template. You can view the selected page or manually choose a different one during the editing process.

Use Cases

- **Password Reset Phish**
  Pair a simulated phishing email posing as a password reset request with a landing page that asks your users to enter sensitive information into a form.

- **Trust but Verify**
  Combine a phishing test masquerading as an internal email with a link to a new policy or other company information with a landing page that asks users to distinguish between their organization's internal emails and phishing emails they have received. These are all branded landing pages that can be adjusted to show what your organization's internal emails look like.

Knowledge Refresher Agent

What It Does

The Knowledge Refresher Agent helps users sharpen their cybersecurity skills and earn badges with short, AI-generated scenario-based quizzes that are a fun and quick way to revisit key takeaways from their recent training.

How To Use It

You can set parameters for knowledge refreshers such as the time interval between training completion and refresher availability. Once activated, users can access their Knowledge Refreshers in their Learner Experience, streamlining engagement with this ongoing element of their learning process.

Use Cases

- **Reinforcing Phishing Awareness**
  AI-generated, scenario-based knowledge refreshers can be used as part of remedial training delivered after a user has clicked on a simulated phishing email to support the initial training. This continuous reinforcement helps maintain vigilance against evolving phishing tactics.

- **Compliance Training Retention**
  By delivering periodic, focused refreshers on key compliance topics, organizations under specific regulatory requirements can demonstrate ongoing due diligence in maintaining a culture of compliance, potentially reducing liability and improving audit outcomes.

Policy Quiz Agent

What It Does

This AI-powered Policy Quiz Agent automatically generates quizzes based on your organization's specific policies, ensuring that employees not only read but also comprehend and retain critical information.

How To Use It

In your KnowBe4 console, you add your organization's policy as a PDF file. After you've added your policy, you can edit and manage quiz questions relating to the policy and then create a training campaign to assign the policy to your users. Once you've assigned the policy to a campaign, you can track your users' progress and acknowledgments.

Use Cases

- **New Employee Onboarding**
  With custom-built policy quizzes, ensure new hires not only review the policies but also demonstrate their understanding. This can provide a measurable benchmark for their initial compliance training and reduce the risk of policy violations due to misunderstanding.

- **Annual Policy Review**
  Streamline your annual policy reviewing process by using the Policy Quiz agent to quickly generate updated quizzes that reflect any changes in company guidelines and rules. Policy-based training campaigns will both help ensure knowledge retention and provide administrators with clear metrics on company-wide policy comprehension, helping to identify areas that may require additional training or clarification.

Deepfake Training Content Agent

What It Does

The Deepfakes training feature allows you to create simulated deepfake videos to further personalize your SAT. By creating a deepfake of a known executive or leader in your organization, you can demonstrate to your users how easily someone can be convincingly impersonated. These videos are locked down and bound to pre-recorded scripts, ensuring the content is used strictly for educational purposes.

How To Use It

KnowBe4 platform users are guided through the creation process in five steps. To ensure deepfakes are only used for educational purposes, KnowBe4 provides a variety of pre-written scripts that mimic the urgency attackers commonly exploit. Based on a submitted video of the deepfake subject, the agent processes the audio to clone the voice. The audio can be regenerated if necessary.

Use Cases

The Deepfake Training Content Agent creates a safe, controlled deepfake training asset that:

- Demonstrates how convincingly attackers can impersonate trusted individuals
- Teaches your employees how to evaluate audio and video cues
- Provides guidance on what to do when something feels "off"

Custom SAPA Agent

What It Does

The Custom SAPA Agent is an intelligent, AIDA-driven evolution of KnowBe4's Security Awareness Proficiency Assessment (SAPA). While traditional assessments provide a trusted baseline, the Custom SAPA Agent transitions organizations toward environment-aware measurement. By understanding an organization's specific security stack and internal policies first, it delivers a tailored assessment that identifies the specific knowledge gaps impacting an organization's unique risk profile.

How To Use It

The Custom SAPA Agent is used to evaluate security awareness, not to train users directly. Admins complete a short Environment Survey so the agent can understand your policies, tools and workflows, then review and launch the assessment. The results provide clear visibility into strengths, weaknesses and risk trends, helping teams plan more targeted security awareness programs and measure progress over time.

Use Cases

- **Environment-Aware Baseline**
  Instead of measuring general knowledge, the assessment aligns to your internal policies and technology, giving leaders a clearer picture of where risk exists. It is recommended to assign the first SAPA after your initial phishing test but before your first training campaign. This establishes a baseline to measure future progress.

- **Data-Driven Planning**
  Use assessment insights to prioritize remediation and shape training campaigns. Because results map to organizational realities, security teams can confidently justify program decisions and accurately communicate risk to leadership.

Conclusion: Why Choose AIDA?

AIDA offers a suite of AI-native agents that transform how you address HRM. AIDA empowers your organization to stay ahead of the curve, using AI to defeat AI threats.

Here's what AIDA brings to you and your organization:

Save Time

By reducing the time it takes admins to create ongoing personalized phishing and SAT for every individual to mere seconds, AIDA frees up your security teams time dramatically.

Align with Industry Standards

AIDA's alignment with the NIST Phish Scale Framework ensures your SAT is consistent with your organization's broader security initiatives.

Unify your Security Teams

Improve relationships, build trust and reduce friction between your security team and other departments by ensuring users are aligned with security objectives.

Turn Efficiency into Budget Savings

By actively managing human risk efficiently, you have flexibility in your security budget to invest in other key initiatives.

KnowBe4 empowers your organization to outpace phishing, vishing, deepfakes and the full spectrum of social engineering. In an environment where humans and AI agents collaborate in real time, your defense must be as fast and adaptive as the threats you face.

See AIDA in Action

Free Phishing Security Test

Find out what percentage of your employees are Phish-prone with your free Phishing Security Test

Free Automated Security Awareness Program

Create a customized Security Awareness Program for your organization

Free Phish Alert Button

Your employees now have a safe way to report phishing attacks with one click

Free Email Exposure Check

Find out which of your users emails are exposed before bad actors do

Free Domain Spoof Test

Find out if hackers can spoof an email address of your own domain

About KnowBe4

KnowBe4 empowers the modern workforce to make smarter security decisions every day. Trusted by more than 70,000 organizations worldwide, KnowBe4 is the pioneer of digital workforce security, securing both AI agents and humans. The KnowBe4 Platform provides attack simulation and training, collaboration security, and agent security powered by AIDA (Artificial Intelligence Defense Agents) and a proprietary Risk Score. The platform leverages 15-years of behavioral data to combat advanced threats including social engineering, prompt injection, and shadow AI.

By securing humans and agents, KnowBe4 leads the industry in workforce trust and defense. More info at **knowbe4.com**.

knowbe4

KnowBe4, Inc. | 33 N Garden Ave, Suite 1200, Clearwater, FL 33755
855-KNOWBE4 (566-9234) | www.KnowBe4.com | Sales@KnowBe4.com