AIDA Positioning

POSITIONING DOCUMENT

AIDA (Artificial Intelligence Defense Agents)

Approved: 12/19/2025

Personas

1. Security Awareness & Program Administrators
Responsible for running phishing and training programs with limited time and resources, often burdened by manual campaign design, tuning, and reporting.

2. Security, Risk, and Compliance Leaders
Accountable for reducing human risk, demonstrating effectiveness to leadership and the board, and aligning awareness programs with broader security strategy.

3. Executive Leadership (CISO, CIO, CSO)
Focused on measurable risk reduction, operational efficiency, and confidence that human risk is being actively managed (not just reported on).

4. End Users (Employees)
Expected to make the right security decisions daily, but often overloaded with generic, forgettable training that feels disconnected from their real work.

The Problem

Your employees are prime targets for today's complex and widespread social engineering attacks. According to the 2025 Verizon Data Breach Investigations Report, up to 68% of cyber attacks involve some form of social engineering, making it the biggest threat for organizations just like yours.

The rise of AI-powered attacks has made this problem even worse. Over 95% of cybersecurity professionals believe AI-generated content makes phishing detection more challenging [LastPass survey 2024]. This technological advancement in the hands of bad actors has created a new breed of highly convincing social engineering attacks that one-size-fits-all security awareness training struggles to combat.

Generic security awareness training is no longer effective in your complex work environment. Trying to run a one-size-fits-all program creates operational burden—leading to manual effort, low user engagement, slow iteration on content, and a limited real-world impact on human risk. Your organization struggles to deliver relevant, engaging content that changes user behavior and resonates across different roles, departments, and even languages. This guesswork and constant tuning are not sustainable.

Specific Problems Target Audiences Face:

1. **Management and InfoSec Leadership:** *"My biggest concerns about AI in cybersecurity are AI-generated phishing, deepfakes, and automated attacks that make threats look real, making it harder for me and my team to detect them. I also worry that AI has become a tool for bad actors, the potential for data leakage, and if AI can protect our network quickly enough."*
    - ○ Struggle to allocate limited security resources for human risk management ○ Find it challenging to quantify and measure the real-world impact of security awareness programs on human risk reduction to executives and the Board. ○ Lack finding the connection and seamless integration between security awareness training and the broader security tech stack
    - ○ Find it difficult to create sustainable, long-term security culture and behavioral change within the organization
2. **Security Awareness Administrators:** *"My biggest concern about AI in cybersecurity is the ability for cyber-criminals to create more believable phishing attacks by removing many of the "tells" that end users were used to queuing [sic: cueing] in."*
    - ○ Struggle to create engaging, personalized content that resonates with diverse user groups
    - ○ Spend excessive time manually crafting phishing templates, aligning difficulty levels and building out campaigns
    - ○ Face challenges in quantifying the real-world impact of their training efforts
    - ○ Struggle to keep training content up-to-date with rapidly evolving threat landscapes
3. **IT and Security Teams:** *"I'm concerned about AI in cybersecurity because it's always easy for criminals to manipulate new technology to better suit their needs early in a product's life cycle. I am unsure if we will be able to respond adequately if we start seeing large scale AI driven cyber attacks against us."*
    - ○ Experience alert fatigue due to the sheer volume of potential threats (an average of 700 social engineering attacks annually per organization [SecureWorld]) ○ Microsoft's 2025 Digital Defense report states that AI-generated phishing attacks are

4.5x more successful than manually created emails.

- ○ Lack the resources to provide timely, personalized feedback on employee security behavior
- ○ Struggle to demonstrate the ROI of security awareness training to leadership

4. Users:

- ○ Suffer from information overload and quickly forget training content that doesn't resonate with them (50% within one day, 90% within one week!)
- ○ Receive generic training that doesn't address their specific role or risk profile
- ○ Lack real-time feedback on their security decisions in day-to-day work ○ View training as a check-the-box item to get through

Your manual, static approaches to security awareness training can't adapt fast enough, creating a major problem: despite investments in generic security awareness training, your organization often fails to see real improvements in human risk. The disconnect between training and real-world application leaves you vulnerable to social engineering attacks, potentially leading to data breaches, financial losses, and reputational damage.

Managing The Problem

AIDA (Artificial Intelligence Defense Agents), is a suite of always-on agents that up-levels your approach to human risk management 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. 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.

How AIDA Works

SmartRisk Agent leverages behavioral data from across KnowBe4's products to help measure human risk. SmartRisk Agent enables AIDA to automatically prioritize risk, target the right users, and adjust interventions - without manual analysis.

Agents work both individually and collectively. With the addition of the **Orchestration Agent**, these agents now work together under an intelligent, goal-driven model that continuously assesses risk and automatically delivers personalized phishing tests and training at the individual

user level—transitioning your organization from manual campaign management to an always-on approach to human risk management.

1. **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. 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.
    a. **Remedial Training Agent:** Automatically assigns the most relevant and engaging content when your users fail a simulated phishing test. This 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.
    b. **Ongoing Training Agent:** Continuously addresses knowledge gaps across your organization and automatically assigns personalized training content to each learner.
    c. **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.
2. **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.
3. **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.
4. **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.
5. **Knowledge Refresher Agent:** AIDA delivers bite-sized Knowledge refreshers at optimal intervals, ensuring your users actually apply critical security concepts.
6. **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.
7. **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.

8. **Human Risk Assessment Agent:** AIDA creates custom risk assessments using your organization's specific documentation and policies, not generic awareness, to assess your users. AIDA selects relevant questions from a large bank to build an org-specific assessment based on your environment.

Key Benefit Statements:

- By reducing the time it takes your admins to create ongoing personalized phishing and security awareness training for every individual to mere seconds, AIDA frees up your security team's time dramatically
- Break away from one-size-fits-all campaigns and provide continuously optimized, individualized security training that adapts to each user's unique risk profile ensuring your organization maintains the lowest possible human risk.
- AIDA's alignment with the NIST Phish Scale Framework ensures your security awareness training is consistent with your organization's broader security initiatives.
- Improve relationships, build trust and reduce friction between your security team and other departments by ensuring users are aligned with security objectives, fostering your organization's security culture and empowering users.

Positioning Statement

For security and compliance leaders at organizations of all sizes who struggle with the operational burden of managing security awareness programs

AIDA is a suite of AI Agents for human risk management

That continuously automates SAT program administration and content personalization so security teams can focus on strategic risk decisions rather than operational tasks.

AIDA doesn't require constant manual configuration and content creation,
AIDA uses autonomous AI agents to handle setup, personalization, and ongoing program management, adapting in real-time to the evolving threat landscape.

Product Description

AIDA is KnowBe4's innovative suite of AI-native security agents designed to automate and enhance human risk management. It provides a comprehensive and adaptive experience that

evolves with the threat landscape. By leveraging multiple AI technologies, AIDA brings together human and artificial intelligence to reduce human risk. AIDA includes advanced reporting by role, executive reports, and more.

AIDA is powered by SmartRisk Agent, a feature of the KnowBe4 Platform that provides a complete view of the human Risk Score within your organization. As AIDA's foundation, SmartRisk Agent leverages behavioral data from across KnowBe4's products to help measure human risk. SmartRisk Agent enables AIDA to automatically prioritize risk, target the right users, and adjust interventions - without manual analysis.

Key features:

- **Orchestration Agent:** Enhances your security program with an "always-on," AI-driven system that automates personalized phishing tests and security awareness training ● **Remedial Training Agent:** Uses AI and hundreds of data points to automatically tailor the most relevant and engaging security training content based on each user's unique risk, needs, and learning style, maximizing retention and real-world application. A sub-agent of Orchestration Agent.
- **Ongoing Training Agent:** Personalized training is automatically assigned to learners to address specific knowledge gaps and reduce organizational risk. A sub-agent of Orchestration Agent.
- **Phishing Agent:** Automated creation and delivery of personalized phishing simulations. A sub-agent of Orchestration Agent.
- **Template Generation Agent:** Leverages generative AI to create highly realistic email templates mirroring current attack vectors while basing social engineering indicators on the NIST Phish Scale Framework
- **Callback Template Generation Agent**: Uses generative AI to create phishing templates specifically for the callback attack vector
- **Knowledge Refresher Agent:** Delivers bite-sized refreshers at optimal intervals ● **Policy Quiz Agent:** Generates intelligent quizzes based on an organization's specific security and compliance policies to ensure users understand and apply critical concepts.
- **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.
- **Deepfake Training Content Agent:** Generates custom deepfake training content featuring a leader from your own organization.
- **Human Risk Assessment Agent:** AIDA creates custom risk assessments using your organization's specific documentation and policies to assess your users.

AIDA is now part of HRM+ SAT.

AIDA Orchestration Feature Write-Up

AIDA Orchestration Feature Write-Up

Positioning:
AIDA Orchestration Agent is an AI-powered agent that handles the security awareness and phishing test administration of your human risk management program. Operating on an "always-on" goal-driven approach, this fully autonomous agent makes decisions independently to create, schedule, and manage both phishing security tests (PSTs) and security awareness training (SAT) at the individual user level. By leveraging KnowBe4's best practices and billions of data points, the Orchestration Agent eliminates the campaign-based approach in favor of continuous, personalized security training that dynamically adapts to each user's risk profile—reducing administrative burden while efficiently driving down organizational risk.

Description:
The Orchestration Agent represents a paradigm shift from manual, campaign-focused security awareness training to an intelligent, fully automated system that manages your human risk management program. This agent acts as a proxy for the traditional administrator, orchestrating the complete lifecycle of phishing simulations and training assignments—from creation and delivery to reporting and optimization.

Key Capabilities:

- **Handles Administration:** Makes decisions independently to determine who to test, what attack vector to use, appropriate difficulty levels, and optimal frequency—all aimed at achieving the lowest possible human risk for your organization.
- **Individual-focused Approach:** Breaks away from large-group campaigns to deliver truly personalized phishing tests and training experiences tailored to each user's specific needs and risk areas.
- **Always-on Operations:** Continuously assesses user risk profiles in real-time, automatically adjusting strategies based on user engagement, performance data, and evolving threat landscapes.
- **Intelligent Orchestration:** Leverages the entire suite of agents (including Template Generation, Remedial Training, Knowledge Refreshers, and Policy Quizzes) to create a cohesive, data-driven security awareness program.
- **Plan-based Control:** Admins maintain strategic oversight by defining constraints called "Plans" that govern how AIDA interacts with specific user groups—controlling factors like phishing and security awareness test frequency while the agent handles tactical execution.
- **NIST Phish Scale Alignment:** Ensures simulations follow industry-standard difficulty frameworks, maintaining consistency with broader security initiatives.

By reducing the time required to create ongoing personalized phishing and training from hours to mere seconds, the Orchestration Agent frees security teams to focus on strategic initiatives while ensuring every individual receives the right training at the right time to effectively reduce organizational risk.