Product Messaging Guide

AIDA (AI Defense Agents)

Autonomous Risk Reduction. Personalized at Scale.

Intended Audience	**Document Purpose**
Sales, Sales Engineers, Channel Partners, Customer Success	Positioning, messaging, talk tracks, ICPs, personas, and competitive guidance

Table of Contents

1. Product Overview & Value Proposition

1.1 Challenges Buyers Face

Manual administration is unsustainable at scale

Security awareness administrators spend enormous time manually building, tuning, and running phishing and training programs. At an enterprise organization with 10,000+ employees, it is simply not possible to build personalized, role-specific training for every individual. The result is one-size-fits-all programs that fail to engage users, fail to change behavior, and fail to reduce real-world risk.

AI-powered attacks have made traditional SAT obsolete

Over 95% of cybersecurity professionals believe AI-generated content makes phishing detection more challenging (LastPass, 2024). AI-powered attacks are 4.5x more successful than manually created emails (Microsoft 2025 Digital Defense Report). Generic training cannot keep pace with highly targeted, AI-generated social engineering at scale.

The disconnect between training and real risk is growing

Despite ongoing investment in security awareness training, organizations consistently fail to see measurable improvements in human risk. The gap between generic, infrequent training and the dynamic, evolving threat landscape leaves organizations vulnerable. Admins lack the visibility, time, and tools to connect training activity to actual risk reduction.

Specific problems by audience:

Management & InfoSec Leadership

- Struggle to allocate limited resources for human risk management
- Cannot quantify or demonstrate the real-world impact of SAT programs to the board
- Lack seamless integration between SAT and the broader security tech stack
- Difficulty creating sustainable, long-term security culture and behavioral change

Security Awareness Administrators

- Spend excessive time manually crafting phishing templates and aligning difficulty levels
- Cannot create engaging, personalized content that resonates across diverse user groups
- Struggle to keep training content current with a rapidly evolving threat landscape
- Cannot quantify the real-world impact of their training efforts

IT and Security Teams

- Experience alert fatigue due to the volume of potential threats (avg. 700 social engineering attacks per org annually — SecureWorld)
- Lack resources to provide timely, personalized feedback on employee security behavior
- Struggle to demonstrate ROI of security awareness training to leadership

End Users

- Forget training content quickly — 50% within one day, 90% within one week

- Receive generic training that does not address their specific role or risk profile
- Lack real-time feedback on daily security decisions
- View training as a checkbox, not a behavior change tool

1.2 The High-Level Pitch (Elevator Pitch)

PITCH	AIDA is KnowBe4's suite of AI Defense Agents — the autonomous orchestration and automation layer of the HRM+ SAT platform. Powered by the SmartRisk Agent, AIDA continuously ingests behavioral data across the KnowBe4 platform and integrated security tools, assigns dynamic individual risk scores, and makes autonomous decisions on how to train each user. The result: security teams stop managing campaigns and start managing outcomes.

For security and compliance leaders who are overwhelmed by the operational burden of running a meaningful security awareness program, AIDA automates SAT administration and content personalization at the individual level — so your team can focus on strategic risk reduction instead of manual program management.

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

1.3 Core Functionality & Agent Suite

AIDA is powered by the SmartRisk Agent, which provides a complete, dynamic view of human risk across the organization. SmartRisk ingests behavioral events from across KnowBe4 products and integrations (SIEM, endpoint, identity platforms) to assign a risk score to individuals, teams, and the organization as a whole. All agents then act on that score.

Agent	What It Does
Orchestration Agent	The always-on engine. Autonomously creates and schedules individualized phishing tests and training campaigns. Admins define parameters; AIDA handles all tactical execution.
Remedial Training Agent	Automatically assigns the most relevant training content when a user fails a simulated phishing test. Sub-agent of Orchestration.
Ongoing Training Agent	Continuously addresses knowledge gaps with personalized content assignments. Sub-agent of Orchestration.
Phishing Agent	Autonomously creates, customizes, and delivers phishing simulations tailored to each user's role, risk profile, and interaction history. Sub-agent of Orchestration.
Template Generation Agent	Uses generative AI to create highly realistic phishing templates mirroring current attack vectors, aligned to the NIST Phish Scale Framework.

Agent	What It Does
Callback Template Generation Agent	Creates phishing templates specifically for callback attack vectors (phishing + vishing combinations).
Knowledge Refresher Agent	Delivers bite-sized refreshers at optimal intervals to combat the forgetting curve.
Policy Quiz Agent	Generates intelligent quizzes based on the organization's specific security and compliance policies.
Recommended Landing Pages Agent	Suggests contextually appropriate landing pages to accompany phishing templates, reinforcing learning objectives.
Deepfake Training Content Agent	Generates custom deepfake training content featuring a leader from the customer's own organization to demonstrate how convincing AI-powered attacks have become.
Human Risk Assessment Agent	Creates custom risk assessments using the organization's own documentation and policies — not generic awareness content.

2. Key Messaging & Strategic Talk Tracks

2.1 Core Strategic Themes

These are the overarching themes to always reinforce regardless of audience.

Theme 1: From Manual Management to Autonomous Risk Reduction

- Value: AIDA eliminates the hours of manual work required to build, personalize, and run security awareness programs. The Orchestration Agent handles what no admin could do at scale — individualized programs for every user, continuously updated.
- Impact: Admins shift from operational work to strategic risk management. Early beta customers reported saving multiple hours per week; some enterprises recovered entire FTE-equivalent capacity.

Theme 2: Real Personalization, Not Just Segmentation

- Value: Generic campaigns group users into buckets. AIDA delivers truly individualized programs — based on each user's actual risk score, learning history, behavioral patterns, and role — using generative AI to create custom phishing simulations and training content at the individual level.
- Impact: Users receive training that is relevant to their specific risks and learning style. Engagement increases, behavior changes, and the organization's risk score decreases.

Theme 3: Outcome-Based, Goal-Driven Program Management

- Value: Admins set risk reduction goals. AIDA works autonomously to achieve them. This is the shift from campaign-based thinking (what are we running this month?) to outcomes-based thinking (what is our risk score and how are we reducing it?).
- Impact: For the first time, security leaders can set a measurable human risk objective and hold the program accountable to it — with data to show the board.

Theme 4: AI That Addresses the AI Threat

- Value: AI-generated phishing attacks are 4.5x more successful than manually created ones. AIDA fights AI with AI — generating hyper-realistic, personalized attack simulations using the same generative AI techniques bad actors use.
- Impact: Users are trained on the exact types of attacks targeting them, not generic threats from three years ago. This keeps your human firewall current.

2.2 Customer Results (Stats & Proof Points)

Lead with customer outcomes and quantifiable results, not features.

Metric	Proof Point
Time savings	Beta customers reported saving hours of manual admin work per week — some eliminated the equivalent of a part-time role.

Metric	Proof Point
Personalization at scale	AIDA delivers individualized programs to thousands of users simultaneously, something no manual process can replicate.
Risk score reduction	Goal-based orchestration continuously drives down organizational human risk scores over time.
Content relevance	Generative AI creates custom phishing templates and training content tailored to each user's role, behavior, and risk profile.
Threat currency	Templates are aligned to the NIST Phish Scale Framework and generated to mirror current attack vectors in real time.
Pipeline performance (CISO whitepaper)	The Top 10 CISO Questions asset tied to the AIDA launch generated nearly $1M in associated pipeline with a 45% conversion rate.

NOTE Always use customer-sourced language in positioning. Buyers trust other buyers. Use beta customer quotes and testimonials to quantify outcomes wherever possible.

2.3 Primary Use Cases

Use Case 1: Enterprise Admin Overwhelmed by Manual Campaign Management

Scenario: A security awareness admin at an enterprise with 8,000+ employees is spending the majority of their week manually building phishing campaigns, selecting content, assigning training, and reviewing reports. Programs are generic — everyone gets the same content on the same schedule regardless of their risk profile or behavior.

AIDA: The Orchestration Agent takes over full program administration. The admin defines parameters (how often, which groups, what constraints) and AIDA handles all tactical decisions — who to test, what attack vector to use, what training to assign, and when. Hours of weekly work become minutes of oversight.

Use Case 2: Security Leader Struggling to Prove ROI to the Board

Scenario: A CISO needs to demonstrate that the organization's security awareness investment is reducing human risk — not just activity metrics. The current program can show completion rates but cannot connect training to measurable risk reduction or demonstrate improvement over time.

AIDA: The SmartRisk Agent provides a dynamic, FICO-like risk score at the individual, team, and org level. Admins set a risk score target and AIDA works to achieve it. Leadership now has a goal-based metric they can track and report on — connecting training activity directly to risk outcomes.

Use Case 3: AI-Powered Attacks Outpacing Traditional Training

Scenario: An organization's existing phishing simulations feel dated. The IT team is seeing AI-generated, highly targeted spear phishing that bears no resemblance to the generic templates being used in simulations. Users are falling for real attacks they were never trained to recognize.

AIDA: The Template Generation Agent and Phishing Agent use generative AI to create phishing simulations that mirror current attack vectors — including deepfakes, callback phishing, and AI-generated spear phishing — at the individual user level. Users are tested on the actual threats targeting them.

Use Case 4: High User Forgetting Rate, Low Engagement

Scenario: Annual or quarterly training is quickly forgotten. Users view it as a compliance checkbox. Completion rates are high but behavior is not changing.

AIDA: The Knowledge Refresher Agent delivers bite-sized reinforcement at optimal intervals to combat the forgetting curve. Training is relevant (personalized to the user's risk), timely (triggered by behavior), and continuous — fundamentally changing the relationship between users and security awareness.

2.4 Strategic Talk Track (Discovery & Demo Guide)

Use this as a guide — not a script. Adapt based on who is in the room.

Stage	Goal	Key Questions to Ask	Core Message to Deliver
Discovery	Uncover the operational burden and the gap between effort and outcomes.	How are you currently building and managing your SAT program? How much time does your team spend on campaign creation and content selection? How personalized is your current program — are different users getting different content? How do you measure whether your program is working?	What we consistently hear is that security awareness teams are drowning in manual work — building generic campaigns that consume hours every week, with limited visibility into whether any of it is actually changing behavior or reducing risk. AIDA was built to solve exactly that.
Value Pitch / Consideration	Establish differentiation. Shift the conversation from features to outcomes.	If your team could get back hours every week from admin work, what would they focus on instead? What would it mean to have a measurable risk score for your organization that you could show leadership? How does your program today address the rise of AI-generated phishing attacks?	AIDA doesn't just automate what you're already doing — it transforms what's possible. For the first time, you can set a risk reduction goal and have an autonomous system work to achieve it. Your admin stops managing campaigns and starts managing outcomes.
Demo / Proof	Show the risk score, Orchestration Agent, and personalization in action.	Can I show you what a SmartRisk score looks like across your org? Would it be helpful to see how a phishing template is generated for a specific user profile? What would it mean to your admin if they could see this level of personalization without doing any of the manual work?	This is what your admin's week looks like with AIDA. They set the goal. AIDA does the rest — testing the right people, with the right content, at the right time, continuously. And every week, your risk score gets closer to where you need it to be.
Decision / Validation	Quantify impact. Support business case for leadership.	What does your current cost look like — time, headcount, program effectiveness? What would a measurable reduction in human risk mean for your board conversation?	Our beta customers quantified exactly what this is worth to them — in hours saved per week, in admin capacity recovered, and in a risk score they can now show the board. I can walk you through those numbers and what they might mean for your organization.

Stage	Goal	Key Questions to Ask	Core Message to Deliver
		What would make this an easy yes for your leadership team?	

3. Business Challenges, Product Alignment & Cross-Sell

Buyer Challenge	AIDA Core Benefit	Cross-Sell / Expand Opportunity
Admin burden: hours of manual campaign management each week	Orchestration Agent fully automates program administration. Admins set goals; AIDA handles execution.	HRM+ SAT Platform: AIDA is the add-on that transforms the existing SAT investment. Sells into any KSAT customer as an upsell.
Generic training that doesn't change behavior	Generative AI creates individualized phishing simulations and training content based on each user's actual risk profile and behavior.	SecurityCoach (SCH): Real-time behavioral coaching based on security stack events (SIEM, endpoint, identity). Extends personalization into the security stack beyond email.
No visibility into human risk or program effectiveness	SmartRisk Agent provides a dynamic, FICO-like risk score at individual, team, and org level. Goal-based orchestration connects activity to outcomes.	PhishER Plus: Transforms user-reported phishing into actionable threat intelligence. Closes the loop between what users report and how the org responds.
AI-powered attacks outpacing traditional simulations	Template Generation Agent uses generative AI to create realistic, current attack simulations — including AI-generated spear phishing, callback phishing, and deepfakes.	KnowBe4 Defend: Inbound email security with behavioral AI. Complements AIDA by defending against threats before they reach users, while AIDA trains users on what gets through.
Proving ROI to leadership and the board	SmartRisk scoring, executive reporting, and goal-based risk reduction give leaders a measurable outcome metric to present to the board.	Compliance Plus: For organizations with regulatory requirements (HIPAA, GDPR, PCI, CMMC), extends the platform to cover compliance training alongside human risk management.

4. Ideal Customer Profiles (ICPs) & Industry-Specific Messaging

4.1 ICPs

Attributes	Ideal (Best Fit)	Moderate Fit	Okay Fit	Poor Fit
Industry	Technology, Financial Services (Banking), Manufacturing, Government	Healthcare, Insurance, Business Services	Education, Utilities, Retail	Small-scale local retail, lifestyle non-profits
Firmographic	SMB: 51-500 employees Mid-Market: 501-1,500 employees Enterprise: 1,501-10,000 employees	Mid-Market: 501-1,500 employees Small: 25-50 employees (may lack alert volume to fully leverage SCH integration)	Small/local firms with few employees wearing multiple hats	Strategic: 10,001+ employees (buying committee complexity)
Persona / Title	IT Manager, IT Director, InfoSec Analyst Security Awareness Administrator	CISO, VP of IT, Security Engineer — focused on high-level risk reduction and board ROI	Compliance Officer, GRC, HR Manager, tech generalist — 'checking the box'	General Counsel, non-technical business owners, administrative roles
Pain Points	Users repeat risky behaviors; lacks resources for personalized programs; high manual admin burden; AI-phishing bypassing traditional training	Hard to prove training effectiveness for audits/insurance/board; high volume of sensitive data; strict regulatory requirements (GLBA, GDPR, CMMC)	Focus on uptime/physical safety; rising insurance premiums; limited bandwidth; IT generalist overwhelmed	Only invests after a material breach; looking for lowest cost seat
Security Team Size	SMB: 1-2 person team (wearing multiple hats) Mid-Market: 3-5 person team Enterprise SOC: 4-10 people who want to automate Level 1 behavioral coaching	Mid-Market: 3-5 person team IT Generalist (1 person)	Entirely outsourced or administrative duty of a single non-security person	No security team
Buying Patterns	Willing to pay to automate; looking to consolidate tools; Q1 primary peak, secondary July-August; multi-year agreements	Longer selling cycles, large buying committees, PoC requirements; healthcare/insurance prefers Q1 and outside open enrollment	Reactive; budget-constrained; restrictive seasonal purchasing; single-year contracts	Extremely reactive; founder/owner gatekeeping; purchasing/holiday freezes

4.2 Industry-Specific Messaging

Industry	Primary Pain Points	How We Win with AIDA
Financial Services / Banking	High-value data makes them prime targets. AI-generated spear phishing, BEC, and deepfake attacks are increasingly common. GLBA and other regulatory requirements demand demonstrable training outcomes.	AIDA's goal-based risk scoring gives compliance teams a measurable outcome to report. Generative AI simulations mirror the exact attack vectors targeting financial employees. SmartRisk connects training to behavioral change over time.
Technology	Centralized infrastructure means one click can cause widespread disruption. Employees are targeted with vendor-impersonation attacks and BEC. Security-aware culture is expected but difficult to sustain.	AIDA's personalization ensures technically sophisticated employees receive appropriately challenging simulations. Automated program management frees lean security teams to focus on threat response, not training administration.
Manufacturing	33% increase in cyber attacks on manufacturing (2024 data). Increasing digitization is expanding the attack surface. Operational disruption from a breach is catastrophic.	AIDA's integration with security stack tools (SIEM, endpoint) allows behavioral risk signals from OT/IT environments to feed the SmartRisk score. Training is personalized to the actual threats targeting manufacturing environments.
Healthcare	Life-safety implications from breaches. HIPAA compliance requirements. Mix of clinical and administrative staff with vastly different risk profiles and training needs. First recorded death linked to a hospital cyberattack was in 2020.	AIDA's individual risk scoring ensures clinical staff, administrators, and IT teams each receive relevant, role-specific training. Compliance reporting supports HIPAA audit requirements. Automation addresses the reality that healthcare security teams are severely understaffed.
Government	Local governments represent ~43% of ransomware victims (Verizon DBIR 2025). Heavy regulations, complex security stacks, compliance documentation burden. Resource and budget constraints are endemic.	PhishER Plus is FedRAMP Moderate authorized. AIDA's automated reporting and risk scoring supports compliance documentation. Personalized, continuous training addresses the challenge of training distributed, mixed-role workforces on a constrained budget.

5. Personas (The Buyers We Talk To)

Role	Pain Points	AIDA Messaging Strategy
Security Awareness Administrator IT/InfoSec Manager	Spends hours every week on manual campaign management. Programs are generic — can't personalize at scale. Can't keep content current with evolving threats. Struggle to show impact or change in user behavior.	Lead with time savings and administrative relief. Demonstrate individualized programs vs. generic campaigns. Show how SmartRisk connects training to behavioral outcomes. Use beta customer quotes on hours saved and admin capacity recovered.
CISO / Security Leader / VP of IT	Cannot quantify human risk or demonstrate SAT program effectiveness to the board. AI-powered attacks are outpacing traditional training. Need to align SAT with broader security stack investments. Board pressure to show measurable risk reduction.	Lead with risk score reduction and measurable outcomes. Connect AIDA to the broader security stack (SIEM, endpoint, identity integrations). Frame AIDA as the tool that makes the security awareness investment defensible. Use the 'FICO score for human risk' analogy — set a goal, track progress, show the board.
IT Director / CIO	Operational burden of managing an enterprise security program. Need to demonstrate ROI and connect security investment to business outcomes. Worried about AI-driven threats making traditional programs obsolete.	Lead with operational efficiency and automation. Quantify what AIDA recovers in admin time and capacity. Connect platform investment to reduced breach risk and insurance cost.
CFO / Finance Leadership	Cost justification for security awareness program investment. Understanding the cost of a breach vs. cost of prevention. ROI of adding AIDA to an existing KSAT investment.	Use breach cost data ($4M average cost per incident). Quantify the cost of manual admin time currently spent on program management. Frame AIDA as a force multiplier on an existing platform investment, not a new line item.
Compliance / Risk Officer / HR	Need to demonstrate training completion and compliance outcomes for audits. Struggle to prove that training is actually changing behavior. Responsible for HIPAA, GDPR, CMMC, or other regulatory requirements.	Lead with compliance reporting, audit trails, and the policy quiz agent. Frame risk score reduction as the new compliance metric. Highlight how AIDA automates the documentation burden alongside the training burden.

6. Competitive Landscape

Messaging note: Do not lead with competitor names unprompted. Use kill questions to surface competitive context, then respond with differentiation. Never position as 'better than X' in external materials — use private landing pages or 1:1 conversations for direct comparisons.

Competitor	Their Weakness	The One Thing to Know (Rebuttal)
Proofpoint Security Awareness	Campaign-based model with limited personalization. Heavy reliance on manual admin configuration. AI features are newer additions to a legacy platform, not native architecture.	AIDA was built as an AI-native orchestration layer from the ground up. The SmartRisk Agent and Orchestration Agent are core architecture, not added features. The result is genuine autonomous program management that Proofpoint's campaign model cannot replicate.
Cofense	Primarily focused on phishing simulation and incident response, not comprehensive human risk management. Limited personalization and no autonomous program management capability.	AIDA goes beyond simulation to autonomous, goal-based risk reduction. The platform connects phishing simulation, training personalization, behavioral coaching, and email security into a unified human risk score — something point solutions cannot deliver.
Mimecast Awareness Training	Awareness training is secondary to their email security core. Platform integration is limited. Personalization and automation capabilities are less mature.	KnowBe4 is the market leader in human risk management with 15+ years of behavioral data powering AIDA's models. Mimecast's training offering lacks the native AI orchestration and risk scoring architecture that makes AIDA different.
Microsoft Defender + Viva Learning / Attack Simulation	Simulation capability is basic. Training content library is limited. No autonomous program management or individual risk scoring. Tightly coupled to Microsoft ecosystem.	AIDA complements Microsoft — it integrates with Microsoft identity and security tools as behavioral inputs to SmartRisk. Where Microsoft offers basic simulation inside the M365 bundle, AIDA delivers a purpose-built, AI-native human risk management platform with a content library of thousands of pieces of training.

Kill Questions (Use in Discovery to Surface Competitive Context)

- How personalized is your current program — does every user get the same content on the same schedule?
- How much time does your admin spend each week on campaign management, content selection, and reporting?
- Can you show your leadership a measurable risk score for your organization today — not activity metrics, but actual human risk?
- How current are your phishing simulations — do they reflect the AI-generated attacks your users are actually facing?

7. FAQs

1. What is the difference between AIDA and the KnowBe4 Security Awareness Training (KSAT) platform?

KSAT is KnowBe4's core security awareness training platform — the content library, simulation tools, and reporting. AIDA is the AI orchestration and automation add-on that transforms how that platform is managed. KSAT provides the content and tools; AIDA autonomously decides how, when, and for whom to use them. AIDA is sold as an add-on to HRM+ SAT.

2. What is HRM+ SAT?

HRM+ SAT is KnowBe4's current platform SKU that includes KSAT (security awareness training, phishing simulations) plus AIDA. It represents the full human risk management platform offering with AI-native orchestration included. AIDA Orchestration is now part of HRM+ SAT.

3. Is AIDA truly autonomous, or does an admin still need to manage it?

AIDA is genuinely autonomous at the tactical level. Admins define 'Plans' — constraints around testing frequency, user groups, and parameters — and AIDA makes all tactical decisions: who to test, what vector to use, what difficulty level, what content to assign, and when. The admin shifts from managing campaigns to managing goals and reviewing outcomes.

4. How is AIDA trained? Where does the data come from?

AIDA is powered by 15+ years of KnowBe4's proprietary behavioral data — the largest dataset of human security behavior in the industry. It was not built on a generic LLM. The models are trained in-house, continuously updated, and maintained with a human in the loop. Customer data is kept on KnowBe4's own infrastructure and is never used to train models for other customers.

5. What are the most common CISO concerns about AI, and how does AIDA address them?

Based on direct research with CISOs at ISSA events during the AIDA launch process, the top questions are:

- Where is the data stored, and is it secure? — All data is maintained on KnowBe4's own servers. Customer data is not shared or used for cross-customer training.
- Is there a human in the loop? — Yes. Admins maintain oversight through Plans, and KnowBe4's team continuously reviews model outputs.
- How is the AI trained? — On 15+ years of proprietary KnowBe4 behavioral data, not a generic LLM.
- How does it handle AI-generated attacks? — The Template Generation Agent uses the same generative AI techniques as attackers to create realistic, current simulations.
- How do we measure if it's working? — SmartRisk provides a dynamic, individual-level risk score. Set a goal; track progress; show the board.

6. How does AIDA integrate with our existing security stack?

AIDA integrates with SIEM platforms (including Splunk), endpoint tools (including CrowdStrike), and identity platforms (including Microsoft) to ingest behavioral events as inputs to the SmartRisk score. This means AIDA is informed by the full security context of the organization — not just what users do inside the KnowBe4 platform.

7. What does 'goal-based' program management mean in practice?

An admin sets a target SmartRisk score for the organization — for example, 'reduce our risk score from 62 to below 45 within six months.' AIDA then makes all tactical decisions to work toward that goal: who needs more frequent testing, which users need remedial training, how to adjust difficulty levels, and when to back off. The admin tracks progress toward the goal rather than managing individual campaigns.