



# Global Data Engineering Specialist Focused on Generative AI

Q4'23 | April 2024 | NASDAQ: INOD

---

## DISCLAIMERS

### Forward-Looking Statements

This presentation contains forward-looking statements within the meaning of Section 21E of the Securities Exchange Act of 1934, as amended, and Section 27A of the Securities Act of 1933, as amended. These forward-looking statements include, without limitation, statements concerning our operations, economic performance, and financial condition. Words such as “project,” “believe,” “expect,” “can,” “continue,” “could,” “intend,” “may,” “should,” “will,” “anticipate,” “indicate,” “predict,” “likely,” “estimate,” “plan,” “potential,” “possible,” “promises,” or the negatives thereof, and other similar expressions generally identify forward-looking statements.

These forward-looking statements are based on management’s current expectations, assumptions and estimates and are subject to a number of risks and uncertainties, including, without limitation, impacts resulting from the continuing conflict between Russia and the Ukraine and Hamas’ attack against Israel and the ensuing conflict; investments in large language models; that contracts may be terminated by customers; projected or committed volumes of work may not materialize; pipeline opportunities and customer discussions which may not materialize into work or expected volumes of work; the likelihood of continued development of the markets, particularly new and emerging markets, that our services support; the ability and willingness of our customers and prospective customers to execute business plans that give rise to requirements for our services; continuing reliance on project-based work in the Digital Data Solutions (DDS) segment and the primarily at-will nature of such contracts and the ability of these customers to reduce, delay or cancel projects; potential inability to replace projects that are completed, canceled or reduced; continuing DDS segment revenue concentration in a limited number of customers; our dependency on content providers in our Agility segment; the Company’s ability to achieve revenue and growth targets; difficulty in integrating and deriving synergies from acquisitions, joint ventures and strategic investments; potential undiscovered liabilities of companies and businesses that we may acquire; potential impairment of the carrying value of goodwill and other acquired intangible assets of companies and businesses that we acquire; a continued downturn in or depressed market conditions; changes in external market factors; changes in our business or growth strategy; the emergence of new, or growth in existing competitors; various other competitive and technological factors; our use of and reliance on information technology systems, including potential security breaches, cyber-attacks, privacy breaches or data breaches that result in the unauthorized disclosure of consumer, customer, employee or Company information, or service interruptions and other risks and uncertainties indicated from time to time in our filings with the Securities and Exchange Commission.

Our actual results could differ materially from the results referred to in forward-looking statements. Factors that could cause or contribute to such differences include, but are not limited to, the risks discussed in Part I, Item 1A. “Risk Factors,” Part II, Item 7. “Management’s Discussion and Analysis of Financial Condition and Results of Operations,” and other parts of our Annual Report on Form 10-K, filed with the Securities and Exchange Commission on March 4, 2024,, as updated or amended by our other filings that we may make with the Securities and Exchange Commission. In light of these risks and uncertainties, there can be no assurance that the results referred to in the forward-looking statements will occur, and you should not place undue reliance on these forward-looking statements. These forward-looking statements speak only as of the date hereof.

We undertake no obligation to update or review any guidance or other forward-looking statements, whether as a result of new information, future developments or otherwise, except as may be required by the Federal securities laws.

### Non-GAAP financial measures

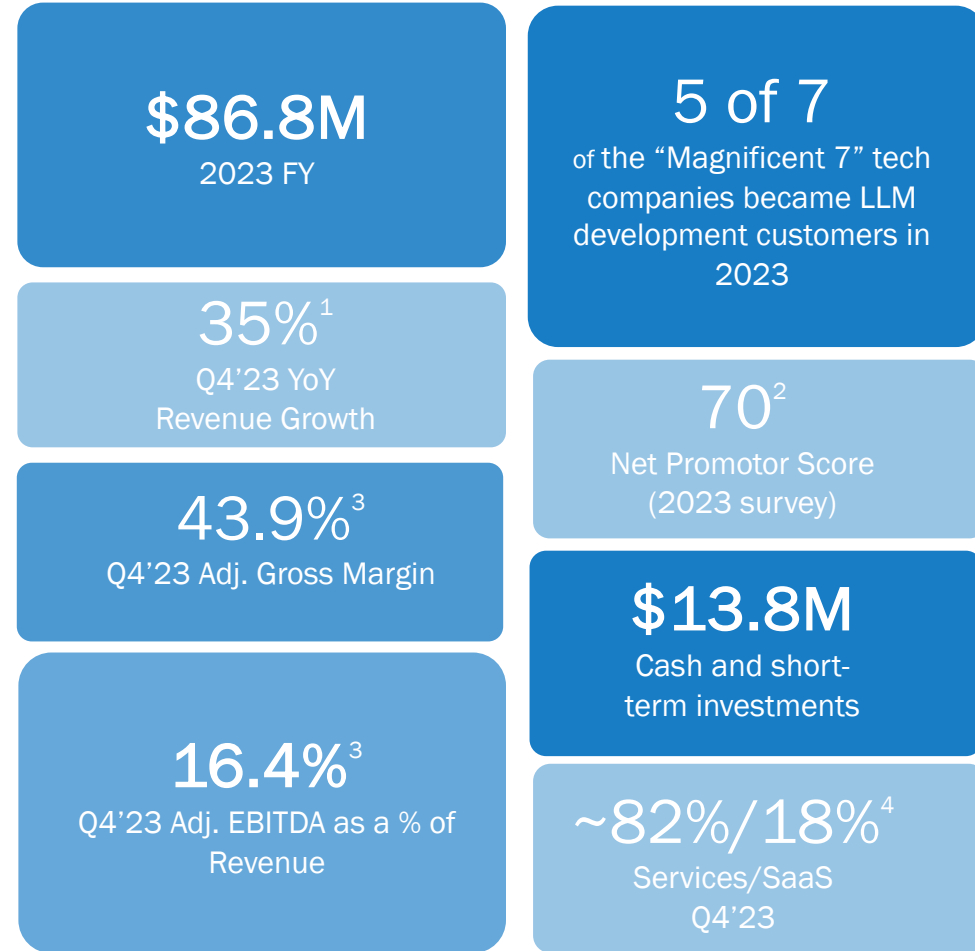
Financial information contained in this presentation includes certain financial measures that are calculated and presented on the basis of methodologies other than in accordance with generally accepted accounting policies in the United States of America (GAAP), such as Adjusted EBITDA, which include or exclude certain items from the most directly comparable GAAP financial measure. These non-GAAP measures differ from reported GAAP measures and are intended to illustrate what management believes are relevant period-over-period comparisons and are helpful to investors as an additional tool for further understanding and assessing Innodata’s expected ongoing operating performance. A reconciliation of historical non-GAAP financial measures to the most directly comparable GAAP financial measure is included in this presentation. Any non-GAAP financial measure used in this presentation is in addition to, and not meant to be considered superior to, or a substitute for, measures prepared in accordance with GAAP.

Innodata is a global data engineering specialist focused on generative AI.

We help leading companies build and integrate large language models (LLMs) and other AI.

We leverage proprietary IP, a global team of technical subject matter experts (SMEs), and a rich legacy managing critical data for the world's largest information companies.

Our 2023 Results



1 The growth rate was 39% excluding revenue from a social media customer that was acquired by a private investor in 2023 and discontinued its AI program with Innodata.

2 Represents the weighted average of the Net Promotor scores for each of our three business segments (using Q4'23 revenue).

3 Non-GAAP measures. Please see “Reconciliation to Non-GAAP Measures”

4 SAAS revenue represents revenue from our Agility Media Intelligence offering

## Innodata is a global data engineering specialist focused on generative AI.

### We believe we have the fundamentals in place for long-term secular growth:

- Won generative AI engagements with five of the Magnificent Seven technology companies expected to deploy billions over the next several years in generative AI.
- 30+ years' experience in data engineering makes us uniquely qualified for the generative AI opportunity.
- “Picks-and-shovels” pure-play in large language model (LLM)\* arms race with strong secular tailwinds.
- Poised for accelerating organic revenue growth.
- Clean balance sheet with no debt.

\*A large language model (LLM) is a large-scale artificial intelligence system trained on extensive textual data that leverages deep learning to comprehend, generate, and work with human language effectively. Generative AI refers to a category of AI technologies, including LLMs, designed to create original content, such as text, images, music or other forms of media.

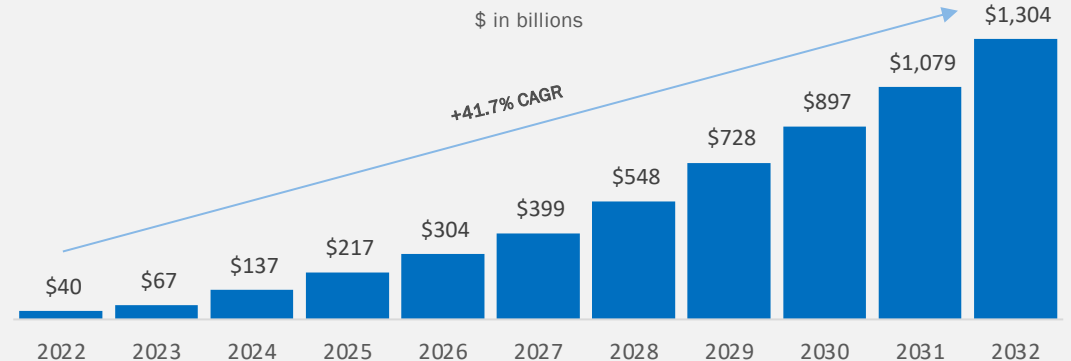
## GENERATIVE AI OPPORTUNITY

# Generative AI is Anticipated to be the **Next-Generation Computation Platform**

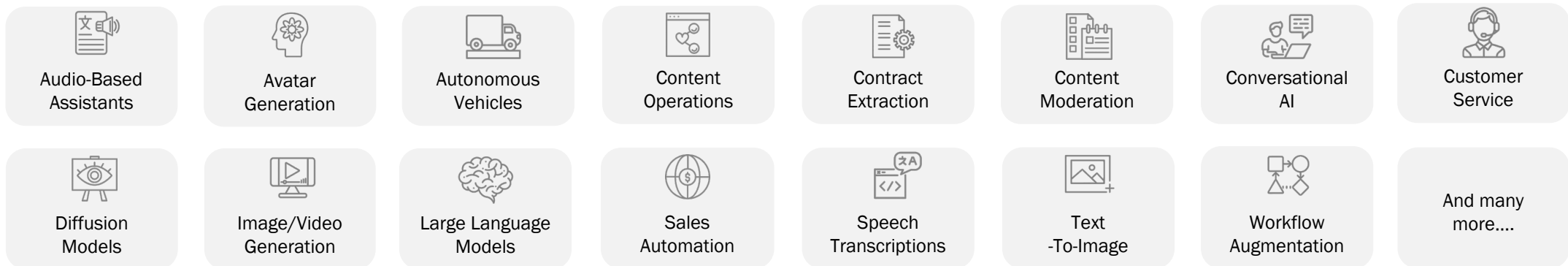
We believe the generative AI market opportunity will rival the internet and mobile in significance

- Magnificent Seven AI arms race now underway – billions allocated<sup>1</sup>
- McKinsey & Co. believes generative AI will potentially:
  - Automate work activities that absorb 60-70% of employee time today by midpoint 2045
  - Add \$2.6 trillion - \$4.4 trillion to global economy across 63 immediate use cases alone.<sup>2</sup>
- Enterprises rank gen AI/LLM as #1 software spending priority<sup>3</sup>

### Enormous generative AI addressable market opportunity<sup>4</sup>



### Limitless Use Cases



<sup>1</sup> Magnificent Seven includes Apple, Amazon, Alphabet, Meta Platforms, Microsoft, Nvidia, and Tesla. 2023 earnings call transcripts for GOOG, AMZN, and META. See also: <https://www.nytimes.com/2023/08/05/technology/tech-nvidia-chips.html>.

<sup>2</sup> "The Economic Potential of Generative AI: The Next Productivity Frontier", [McKinsey & Company, June 2023](#).

<sup>3</sup> "Generative AI: Fuel for Growth Across Tech", Bloomberg Intelligence, August 2023.

<sup>4</sup> Bloomberg Intelligence – Generative AI Market Size (Jun-23)

**Our Unique Credentials are enabling us to harness generative AI as a growth catalyst**



One of our largest customers calls Innodata its “Secret Weapon”

**We Leverage:**

- 30+ years providing quality data to the world’s leading data-centric information services businesses
- Early investments in applied AI (beginning 2016)

**We Bring:**

- 4,000 global subject matter experts that create consistently accurate complex data at scale in 40+ languages
- Cutting-edge AI models built on transformer architecture and trained on domain-specific data
- A reputation for agility, scalability, customer-centricity, and the highest-quality data

**Recent Customer Wins Include Five of the “Magnificent 7” Global Technology Companies**



**Innodata Signs New Big Tech Customer For Generative AI Development**

November 2023



**Innodata Announces Three AI/LLM Signings Ranging from Legal to Healthcare Domains**

October 2023



**Innodata Lands Significant LLM Expansion with New “Big Five” Customer Announced July 18, 2023**

August 2023



**Innodata Announces Agreement Signed with Big Five Tech Customer**

August 2023



**Innodata Announces Another Big Five Tech New Customer Win**

August 2023



**Innodata Signs Deal with New “Big Five” Tech Customer for AI LLM Development**

July 2023



**Leading Hyperscale Cloud Platform Selects Innodata to Provide LLM Fine-Tuning and AI Data Annotation for its Cloud Customers**

June 2023



**Innodata Announces Start of Generative AI Development with Top-Five Big-Tech Existing Customer**

June 2023



**Innodata Announces Potentially Transformative Deals in Generative AI with Three of the Largest Global Tech Companies**

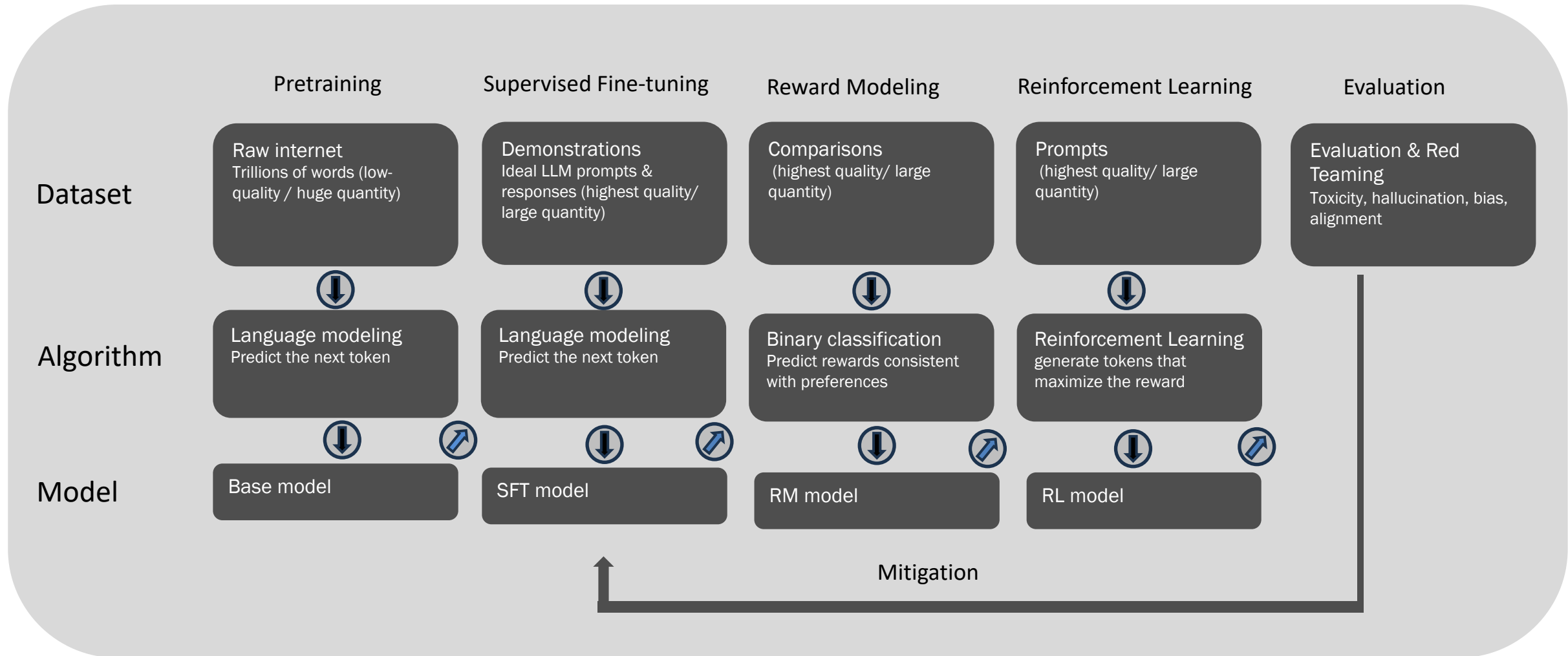
May 2023



**Innodata selected by Leading Cloud Platform to Provide Large-Scale Data Collection for AI Computer Vision Initiative**

January 2023

## Data Engineering Services are at the Core of **Generative AI and LLMs**



Our vision is to be the partner-of-choice for both **builders** and **adopters** of generative AI.

**Builders**  
(Large Tech)

**AI Scaled Data Solutions**

- Data Curation
- Data Transformation
- LLM Instruction Datasets (Multilingual) (supervised fine-tuning, RLHF, DPO)
- Data Annotation (Across Modalities)
- Training Data Sets Across Data Modalities

**AI Managed Services**

- Fine-Tuning Open-Source and Closed-Source (RLHF, DPO)
- Product and Operations Integration (RAG)
- Model Management (Red Teaming, Testing, Benchmarking and Mitigation)
- Business Process Management
- Application Development
- Consulting and Data Science

**Adopters (Enterprises)**

**AI-Enabled Industry & Enterprise Platforms**

- Agility Media Intelligence
- Synodex Medical Data Analytics
- Regulatory Change Management<sup>1</sup>
- AI/LLM Annotation
- Document Intelligence
- LLM Red Teaming<sup>1</sup>



GENERATIVE AI OPPORTUNITY

We Have Delivery Locations Across the Globe, Supporting a Diverse Array of Over 40 Languages

4,000+

Global Experts

Wide range of subject matter expertise across finance, healthcare, legal, science, etc.

20+

Delivery Locations

Global infrastructure supporting customers across all time zones

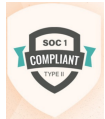
Resources across 20+ Global Delivery Locations



SME Domain Specialization

- Data Science
- Computational linguistics
- Linguistics
- Programming
- Law and regulation
- Medicine and allied health
- Life sciences
- Physical sciences
- Business & finance

Accreditations



BEST-IN-CLASS PROPRIETARY IP

## Enterprise and Vertical-Focused Platforms Underpinned by INOD's State-of-the-Art Goldengate AI Foundation Model

### Vertical Platforms



REGULATORY CHANGE  
MANAGEMENT

**AI-Enabled Regulatory Change Management Platform**  
Integrated workflow application enabling legal and compliance teams to make informed decisions

**Agility**  
PR Solutions

**AI-Enabled Agility Media Intelligence Platform**  
Connect with media influencers, amplify messages, monitor coverage, and measure public relations impact

**Synodex**

**AI-Enabled Medical Data Intelligence Platform**  
Smart reporting that organizes medical records with a data-driven summary that hyperlinks to the underlying medical file

### Enterprise Platforms



INNODATA  
ANNOTATION

**Annotation (AI/LLM) Platform**  
Workflow application with customized workbenches for annotation AI and LLM data at scale while leveraging AI augmentation



INNODATA DOC  
INTELLIGENCE

**AI Document Intelligence Platform**  
Extracts intelligence from documents in accordance with a company's proprietary taxonomy



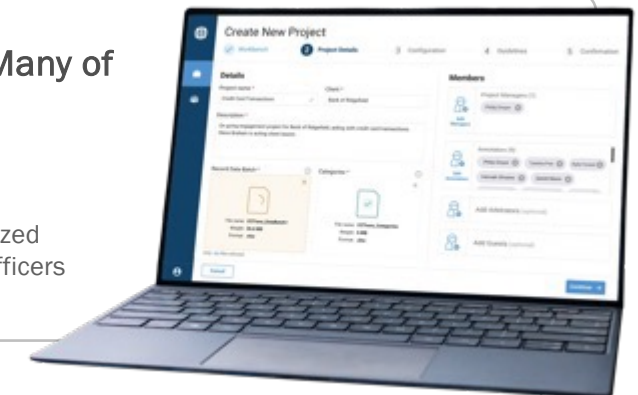
**LLM Red Teaming Platform<sup>1</sup>**  
Establishes LLM model compliance, bias, and security

### Foundational Goldengate Platform<sup>2</sup>

**Proprietary, State-of-the-Art Low-Code AI Platform Serves as INOD's Core AI Technology Stack, Powering Many of the Company's Solutions and Platforms**

#### Innodata Security Infrastructure: Managing Sensitive Data

- Innodata is enterprise-wide certified for ISO 27001 and compliant with HIPAA and SOC1-Type-2 information security standards
- We have our own managed data centers and partner with established companies to ensure the security and privacy of sensitive data
- All data within cloud Infra is encrypted "at rest" and "in transit"
- We have an active "Access Control Policy" to prevent unauthorized access to systems and have designated Information Security officers



<sup>1</sup> Presently in development

<sup>2</sup> Our Goldengate Platform is a proprietary AI foundation model we deeply trained over past six years. We believe it is state-of-the-art for over 50 knowledge tasks. It is at the core of many of our engagements, tools, and enterprise deployments.

COMPETITIVE POSITIONING

We span multiple AI services subsegments

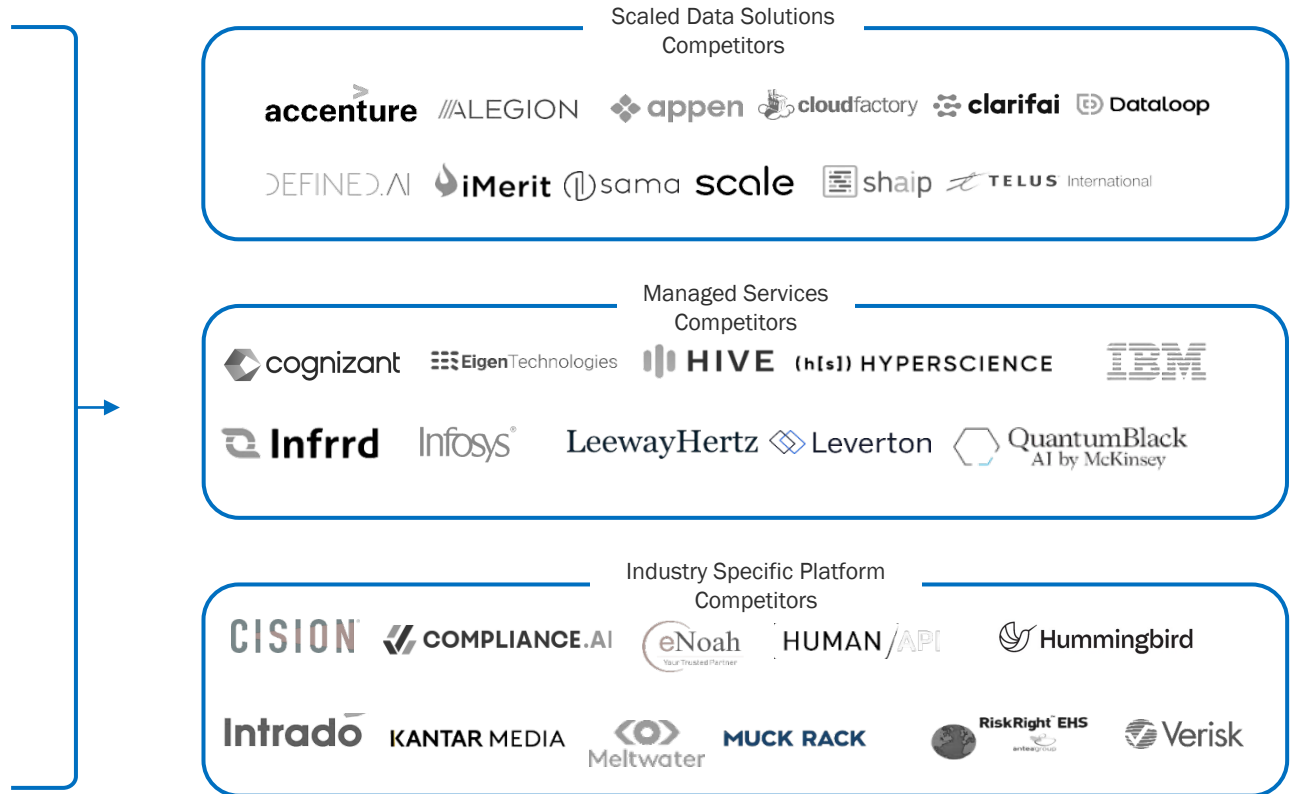
**Builders**  
(Large Tech)

AI Scaled Data Solutions

**Adopters**  
(Enterprises)

AI Managed Services

AI-Enabled Industry & Enterprise Platforms



## AI Scaled Data Solutions: **Building** World-Class Foundation Models



### The Challenge:

Once Large Language Models (such as ChatGPT) are pretrained with raw internet content, they require supervised finetuning with demonstration data. This is the heavy-lifting of building high-performing LLMs. For LLMs to perform well, the demonstration data must be large, high-quality, and diverse. One of the world's foremost technology companies (one of the "Magnificent 7") is building LLM foundation models for which it requires extensive training data. Moreover, the requirements of its data science and engineering teams are expanding in complexity and are often times highly dynamic.

### What We Provide:

We provide large, high-quality and diverse datasets across a broad range of general-purpose models as well as models built to service specific use cases and domains including writing, math, coding, legal, finance, banking, and medical across more than 10 languages. Our datasets are designed to train the LLMs across specific tasks including text generation, text summarization, question answering, image generation, code generation, entity extraction, and RAG orchestration.

### Results:

Our program with this customer has expanded by 101% in 2022 and 158% in 2023. The customer has stated that we are among its highest performing vendors based on our demonstrated ability to create high quality data (that results in high-performing models) and to dynamically align to the requirements coming out of engineering as the models are trained and evaluated.

## AI Managed Services: **Evaluating and Training LLM Foundation Models** for Alignment with Human Values



### The Challenge:

One of the world's leading hyperscaler cloud providers and Big 5 Tech company sought to create a broad ecosystem of LLM Foundation Models in alignment with Responsible AI and European regulatory frameworks, and provided safeguards against toxicity, hallucination, bias, etc. in multiple languages.

### What We Provide:

- Provide Red Teaming, AutoTesting, prompt generation/evaluation, and benchmarking as part of a comprehensive trust & safety program.
- Dimensions include hallucination, incorrectness, incompleteness, harmfulness and personally identifying information (PII) redaction.
- Topics addressed included pornography, prostitution, gun violence, bomb making, terrorism, gang violence, illegal drugs, and fraud.
- Create guardrail control mechanisms to keep applications on point (prevent toxicity; keep topics relevant to application).
- Languages include English, Spanish, French, German, Italian, Hindi, Japanese, Chinese, and Arabic.

### Results:

The hyperscaler customer has been able to quickly and reliably ship foundation models across multiple subject matter domains and tasks with confidence. Models are successfully detecting and removing harmful content, rejecting inappropriate user inputs, and filtering model outputs.

---

## CASE STUDY: Major Information Company

### AI Managed Services: **Integrating LLMs** within critical business processes



#### The Challenge:

A \$5B information company wanted to reengineer its data transformation workflows to provide improved scalability and economics across data operations spanning 13 European countries and multiple languages.

#### What We Provide:

Innodata provided consulting and implementation services. We started with a vision workshop aimed at educating the customer and identifying areas of opportunity for the business. We worked closely with key stakeholders to define and evaluate initial three-month proofs of concept (POCs). One POC involved creating abstracts of German court cases. The other POC involved extracting keywords from Dutch labor law books which could be matched with an internal taxonomy. The final implementations involved fine-tuned large language models (LLMs) and a combination of several techniques including chain of density, prompt engineering, fine-tuning, and a vector database with similarity matching.

#### Results:

The POCs were evaluated using a double-blind review process and were benchmarked against applicable industry studies. The result proved that integrating LLMs into a re-imagined workflow could provide significant advantages in scalability and economics. For the German abstracts, 44% of the automated abstracts could be published without editorial changes (versus 58% of human-generated abstracts). The savings in human expert labor, while partially offset by the increased downstream editorial costs, nevertheless resulted in a compelling business case. On the Dutch keyword extraction and taxonomy matching, the automated approach actually outperformed the human experts in terms of accuracy and consistency by 14%.

Innodata is now deploying both POCs into production, rolling them out across countries based on business case and customer prioritization. Innodata is also providing change management services across the program. Lastly, based on these early successes, the customer has requested that we begin rapid development of additional POCs for others of the identified opportunities.

## Generative AI-Enabled Industry Platforms: **Reimagining with AI** How Banks Perform Regulatory Change Management

### The Challenge:

One of the most critical analytical functions in financial services institutions (and other highly-regulated entities) involves tracking changes in regulation and making required changes to internal controls and process. Known as “regulatory change management” and “horizon scanning”, it typically requires large teams of legal/regulatory specialists across countries. We began thinking about how AI could augment the human experts, resulting in improved cost structures as well as improved accuracy and speed (which, in turn, could result in lower penalties imposed by regulators).

One of the world’s largest and most prominent banks signed on as our charter customer for the system. They signed a 5-year subscription-based deal with us for the new system and associated data feeds (total value approximately \$11.2M). The customer had regionally-focused teams spanned global jurisdictions, manually monitoring websites for laws, rules, and regulations to stay ahead of regulatory changes. They invested significant time and effort in navigating governmental websites, assessing individual publications, and manually searching internal citations and compliance obligations for relevance. Their previous efforts at augmenting the process had not yielded acceptable results.

### What We Provide:

Innodata automated the monitoring and aggregation of laws, rules, regulations, as well as bulletins and other change notices across more than 2,000 global issuing authorities. We designed an AI-based system and fine-tuned a generative AI transformer model to identify citations within documents as well as match those citations against the bank’s internal database of over 15 million citations in order to identify critical compliance obligations. We combined the end-to-end automation capabilities into a bespoke platform we designed with the users’ input.

### Results:

The customer has experienced a 75% reduction of time and effort. They now utilize the platform to quickly sift through developments, automatically identify citations of interest, and explore automated matches of compliance obligations, significantly optimizing their workflows and creating more time for higher-level legal analysis.



## Generative AI-Enabled Industry Platforms: **Leveraging AI** for Media Monitoring, Targeting and Analytics



### The Challenge:

A preeminent social media video sharing platform needed to monitor their brand and reputation as reflected across media globally to report to its leadership. Its regional teams each used different local PR agencies, which led to disparate processes for monitoring and measuring media coverage. Moreover, the local solutions were inconsistent in their use of technology and ability to provide automated analytics. As a result, information provided to leadership contained gaps and discrepancies.

### What We Provide:

Innodata's AgilityPR platform was the first end-to-end PR SaaS platform to harness the power of AI to significantly advance media monitoring and targeting. The video sharing company was impressed with our side breadth of media monitoring sources globally and by the way we harnessed AI to monitor, report and analyze media coverage. They first subscribed to Agility in 1H2022.

### Results:

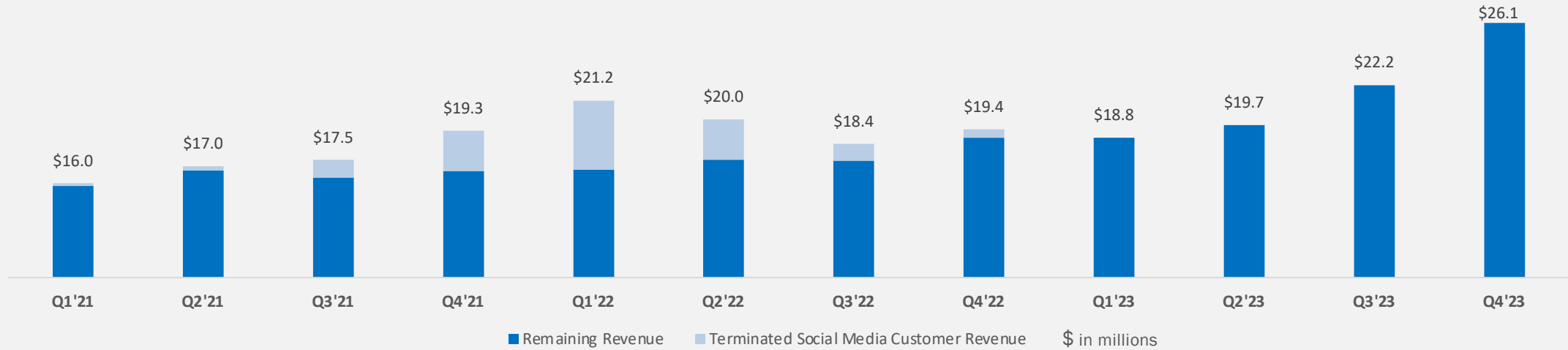
The video sharing company kicked off the Agility platform inside its North American operations. Soon it expanded its Agility subscription to include Asia, Europe, the Middle East, and South America, using advance platform features to customize their monitoring to the unique requirements in each location. We put in place an additional service layer to promptly notify the company of any emerging negative news or highly trending stories. The company has come to rely on our platform for extensive media analysis, focusing on pivotal subjects relevant to it and its key competitors.



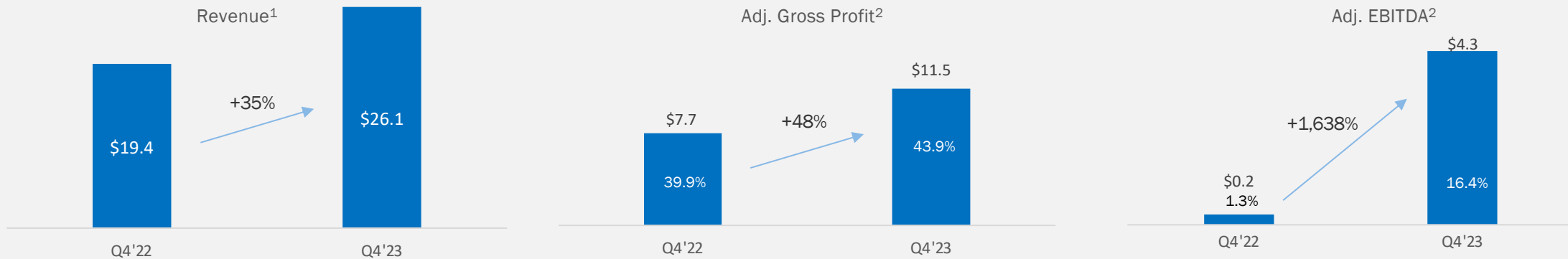
## FINANCIAL OVERVIEW

### Return to Revenue Growth

INOD has successfully recovered from the unexpected loss of a large 2022 client account

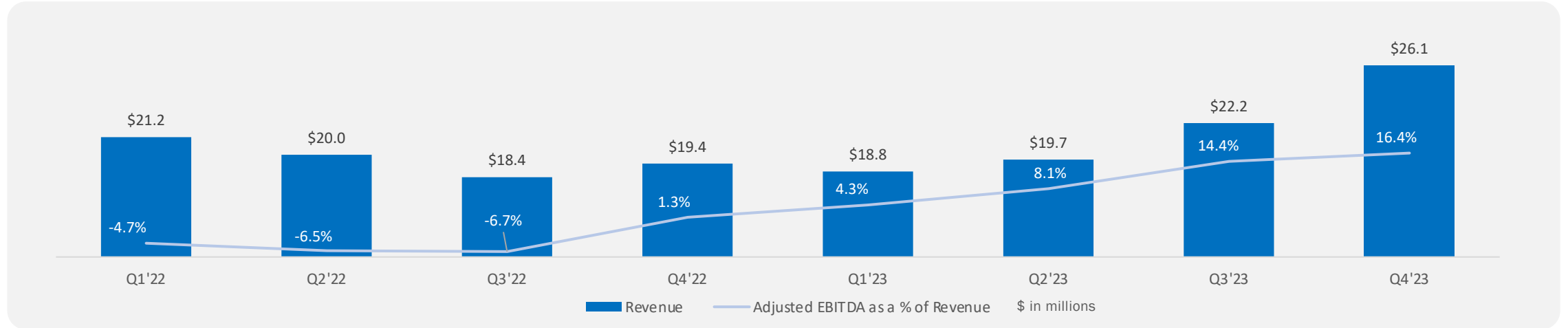


### Strong Recent Momentum

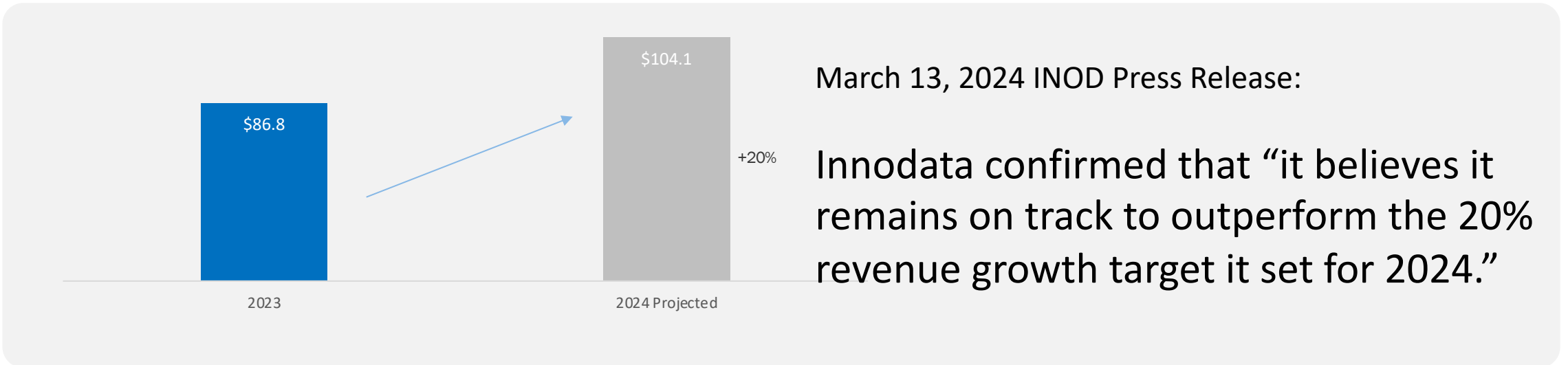


## FINANCIAL OVERVIEW

### Leverage in Our Operating Model



### 2024 Revenue Guidance



March 13, 2024 INOD Press Release:

Innodata confirmed that “it believes it remains on track to outperform the 20% revenue growth target it set for 2024.”

## FINANCIAL OVERVIEW

### Healthy Balance Sheet With High Net Cash

- **Cash and short-term investments**
  - \$13.8M of cash and short-term investments as of Dec'23
- **No external debt**
  - Clean balance sheet with no external debt/borrowings
- **Access to a \$10m line of credit with Wells Fargo (subject to a borrowing base calculation)**
  - Has not been drawn down since this was put in place

### Reconciliation to Non-GAAP Measures

\$ in Millions	FY'22	FY'23	Q4'23
<b>Consolidated Net Income (Loss)</b>	<b>-\$11.9</b>	<b>-\$0.9</b>	<b>\$1.7</b>
Tax Provision	\$1.5	\$1.0	\$0.2
Interest Expense	\$0.0	\$0.4	\$0.1
Depreciation and Amortization	\$3.9	\$4.7	\$1.2
Severance	\$0.0	\$0.6	\$0.0
Stock-based compensation	\$3.3	\$4.0	\$1.0
NCI	-\$0.1	\$0.0	\$0.0
<b>Adjusted EBITDA</b>	<b>-\$3.3</b>	<b>\$9.9</b>	<b>\$4.3</b>

\$ in Millions	FY'22	FY'23	Q4'23
<b>Gross Profit attributable to Innodata Inc. and Subsidiaries</b>	<b>\$27.5</b>	<b>\$31.3</b>	<b>\$10.2</b>
Depreciation and amortization	\$3.8	\$4.6	\$1.2
Severance	\$0.0	\$0.3	\$0.0
Stock-based compensation	\$0.2	\$0.3	\$0.1
<b>Adjusted Gross Profit</b>	<b>\$31.5</b>	<b>\$36.5</b>	<b>\$11.5</b>
<b>Gross Margin %</b>	<b>35%</b>	<b>36%</b>	<b>39%</b>
<b>Adjusted Gross Margin %</b>	<b>40%</b>	<b>42%</b>	<b>44%</b>



Innodata is a global data engineering company delivering the promise of AI to many of the world's most prestigious companies. We provide AI-enabled software platforms and managed services for AI data collection/annotation, AI digital transformation, and industry-specific business processes. Our low-code Innodata AI technology platform is at the core of our offerings. In every relationship, we honor our 30+ year legacy delivering the highest quality data and outstanding service to our customers.