TidyAI: Automating creating Tidy Data

Automate the Grunt, Amplify the Insight

tidyai

The Detriments of Dirty Data

Dirty data is a significant hurdle in the data analytics process: “Garbage in, garbage out”

Compromised Decisions

Any mistake leads to model inaccuracies, misleading advice, and in the end to flawed decisions. Companies risk making expensive mistakes.

Missed Steps

The tedious nature of data cleaning leads data scientists to skip essential preprocessing steps.

Error-prone

Manual data cleaning is prone to human errors. Over-cleaning can lead to the loss of vital information, while under-cleaning can leave inconsistencies.

Time-Intensive

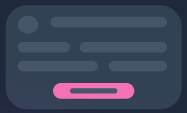
Data cleaning is the most time-consuming task often taking 50-80% of the total project time.

Today's Data Cleaning: Manual, Time-Consuming, Inefficient

Data scientists spend hours working in sheets or notebooks iteratively cleaning the data and checking the results.

Some automation tools are available but they cannot adapt to the specific data or require a specification of the wanted transformation and target columns.

- Manual Data Review to Find Errors
- Long winding iterative approach
- Limited automation
- Non-adaptive to the data or analysis requirements



The technology is ready, AI is getting accepted for decision making

Social

- AI acceptance: With ChatGPT , an AI tool has been accepted broadly for the first time.
- Changing Role of Data Scientists: Data scientists are too expensive and seen as strategic partners. They are now needed for high-level analysis and insight generation rather than grunt work.

Economic

- Rising Volumes of data: data in business is growing exponentially. Cleaning this data manually is becoming financially unsustainable.
- Cost of mistakes: As data-driven decision-making becomes the norm, the cost of decisions based on dirty data is skyrocketing.

Technology

- AI adaptability: SOTA models can adapt to new scenarios and challenges. It can learn and adjust its strategies, making it perfect for diverse and ever-evolving datasets
- Planning capabilities: Today's AI can plan and reflect to take over tasks that were traditional in the human domain.

An intelligent agent tidying your data

Data cleaning isn't a static task but a dynamic journey towards tidy data.

AI-Powered Planning

Our AI intelligently inspects datasets, pinpointing anomalies and potential issues. It doesn't just recognize errors; it understands the context, ensuring data retains its meaning post-transformation.

Automated Transformations:

Once the planning phase identifies necessary transformations, the agent seamlessly executes them. This automation not only reduces manual intervention but ensures consistent, error-free transformations every single time.

- Manual data review to find errors
- Non-adaptive to the data or analysis
- Long winding iterative approach
- Limited automation

Iterative Reflection & Refinement

Post transformation, it isn't done. The AI checks the modified dataset to identify any further cleaning or transformation needs. This iterative approach guarantees a level of data cleanliness and structure that's hard to achieve with a one-and-done solution.

- Error-prone
- Missed Steps
- Time-intensive
- Compromised Decisions

Data Scientists and Analysts can focus on what makes them great

Data Scientists	Business Analyst	Data Engineers	Executives
Focus on core analytics and modeling.	Derive insights from pristine data without the technical hurdles.	Seamless integration and data pipeline management.	Make strategic decisions with the assurance of clean, reliable data.
• Eliminate Time-intensive Cleaning • Avoid Error-prone Manual Interventions	• Confidence in decisions • Streamlined analysis workflow	• Efficient Data Transformation • Mitigate Missed Transformation Steps	• Drive Data-Backed Strategies • Enhance Business Agility with Clean Data

Competition

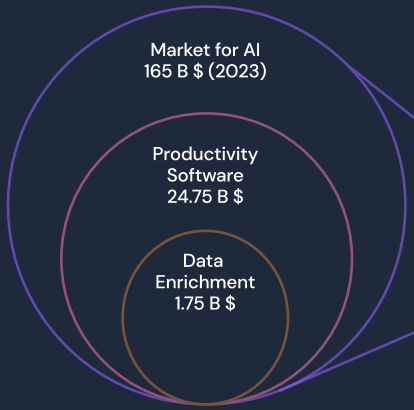
Data Cleaning is the missing tool to empower data scientists

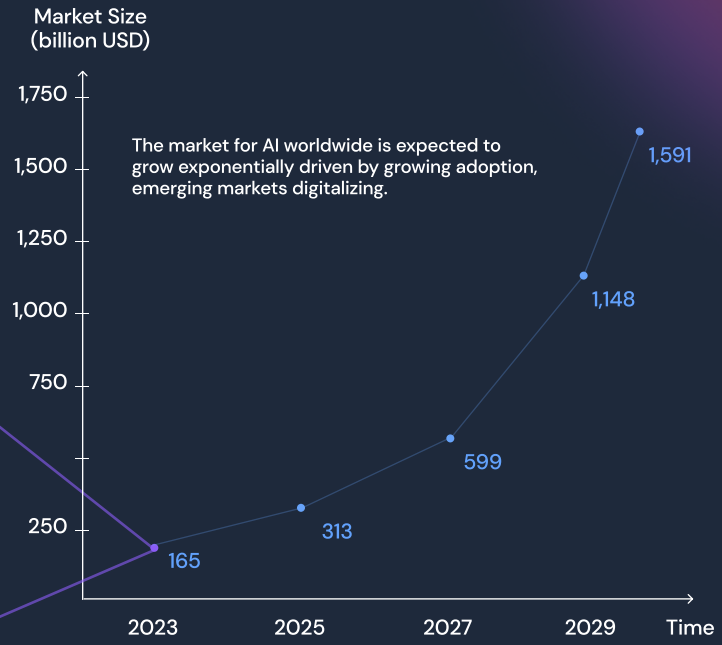
There are many challenges to tackle in data cleaning.
We don't just address issues—we resolve them

	tidyai	Akkio	WinPure™	OpenRefine
Reduced Manual Labor	AI controlled transformations	AI transforms on clean data	Transformation has to be chosen	Transformation has to be chosen
Speed & Efficiency	Full spectrum automation	No automation for cleaning	No automation for cleaning	No automation for cleaning
Beyond Static Solutions	Model learns from errors	Doesn't capture user data	No AI / user feedback	No AI / user feedback
Precision in Action	Quality control integrated	QC through limitated agency	QC through pre-definition	QC through pre-definition

Digital Robots are expected to grow faster than the overall market

The global market for digital robots is roughly 12.75 B $ and for data enrichment solutions 2.125 B $. This can be expected to grow rapidly with data volumes in enterprises growing 12% a year and the lack of talent in data engineering.

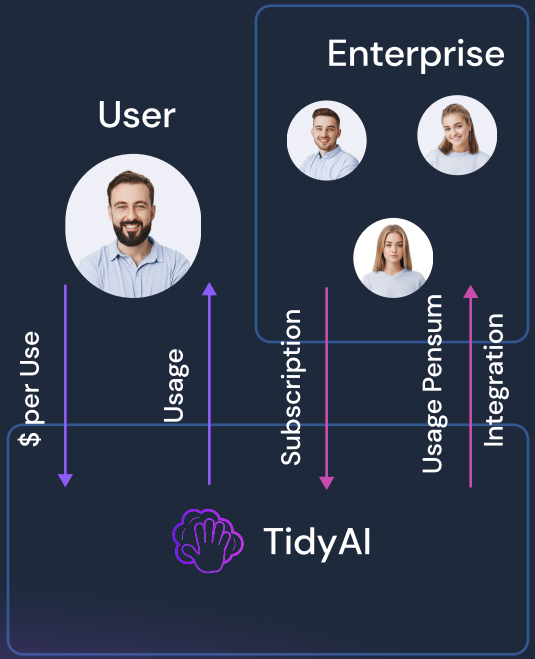




Usage-based sales to enable a product-led sales model

We want to align our business model with the value our customers receive. We charge based on the number of transformations applied to the data.

We sell to data scientists first to prove value and aim to upsell to the enterprise over time by selling a subscription with a certain usage pensum and a custom integration (white glove) into the enterprise's system.



We do not want to be a product but build an ecosystem of tools and users

We want to build an ecosystem, where data scientists can collaborate and improve the AI. To add custom functionality, we will enable our users to build custom transformations that they can open-source to other users to sell on a marketplace and enterprises to build their own transformation suites that can be executed on new data.

This is supported by our agents, which take care of cleaning the data, but also understand the business context and learn to perform analysis over time.

This ecosystem will allow us to capture valuable data to improve our models over time.

Reshaping the data narrative

From our humble beginnings focusing on data cleansing to our ultimate vision of revolutionizing the role of data scientists

🔎 Initial Launch

Launch the initial version of our agent to give data scientists superpowers in the tedious task of data cleansing.

End Q3 23

🎨 Data Cleaning 2.0

Introduce advanced cleansing techniques, anomaly detection, and quality control to be the gold standard.

Q4 23

📦 Integration

Seamlessly integrate into popular data tools and platforms to embed the agent into the core ETL pipeline.

Q1 24

💻 Analytical Bots

Deploy AI bots that can take high-level instructions from data scientists to perform analysis on clean data.

Q2-3 24

Meet our amazing team

We combine a unique background of AI, backend, frontend, and product, all with a history in AI and data, so we know the pains of our customers.

Mikhail Azaryan
Digital Wizard

Nicolay Gerold
Product Puppeteer

Robert Lukoshko
Neuron Nurturer

Alex Pokras
Server Whisperer

David Podolskyi
Pixel Perfectionist

Let’s elevate data cleaning from a chore to a strategic advantage. Dive into the future with us.

Contact us at:

Appendix

Detailed Market Study

tidyai

Resources

1. State of AI in 2022, Mckinsey
2. Why Data Scientists aren't Data Engineer, Forbes
3. Market for Digital Robots, Statista
4. Data Enrichment Market Size, Global Newswire
5. Worldwide Data Created, Statista
6. AI Market Size, Statista