Data Science Group A Global AI COE

Dr. Elan Sasson – CEO
Dr. Gideon Rosenthal – R&D

DATA
SCIENCE
GROUP

About DSG

- Founded six years ago, DSG is a profitable bootstrap company

- DSG consists of two main divisions

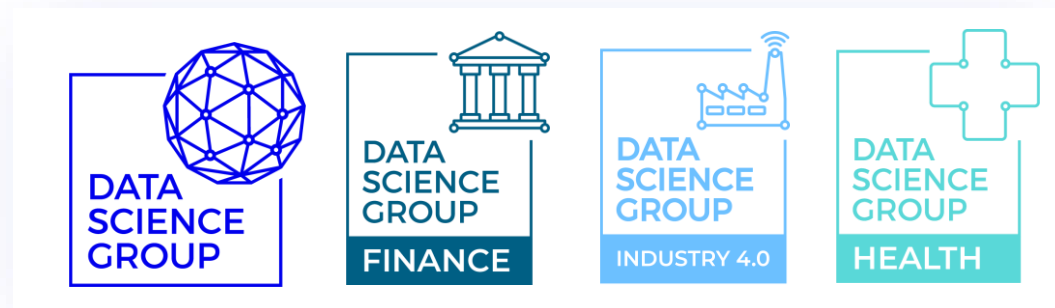**AI-enabled business solutions** ⟷ **AI Products & Platforms** 🧠 **e-volve** 💗 **MD**EEP.AI

- Team: 30 data scientists

- Dozens of projects with a success rate of over 96%

- Unique cross-vertical knowledge in solving complex AI problems

- Access to variety of data sets in diverse domains, different scales and formats of data

- Global Presence

- Patent Pending

DSG Leadership

Dr. Elan Sasson
CEO

DSG Co-founder. Serial entrepreneur, board member, and Lecturer. Member of LAMBDA AI Lab. Expert in Business Intelligence, data science, machine learning, and mining techniques.

Dr. Gideon Rosenthal
Head of Research

Data science, machine learning, deep learning, graph theory, computational neuroscience, network analysis, statistics, and big data technologies.

Dr. Amjad Abu-Rmileh
Chief Data Scientist

Machine (deep) learning, signal processing, time series analysis, brain computer interface, recommendation systems, modelling and model-based control algorithms.

Dr. Gal Noyman-Veksler
VP Business Development

Business executive and entrepreneur. Behavioral researcher (Wolf Award winner). Past roles in venture capital firms in healthcare and startups.

Dr. Danielle Afterman
Head Statistician

Statistics. Mathematics. Machine learning and Big data modeling in relation to statistical models. Hidden Markov Models. Operation research with emphasis on data analysis and insights derivation.

Dr. Orna Berry
Advisory Board

Former chief scientist and head of the industrial R&D operations of the Israeli Ministry of Industry, Trade and labor. Entrepreneur, investor, and board member. Past roles included Dell EMC GM, Israel Center of Excellence.

Prof. Carmel Sofer
Advisory Board

Research fellow in computational psychology. Entrepreneur, investor, and active board member. Past roles included President of Comverse Europe.

Use Cases - Sample
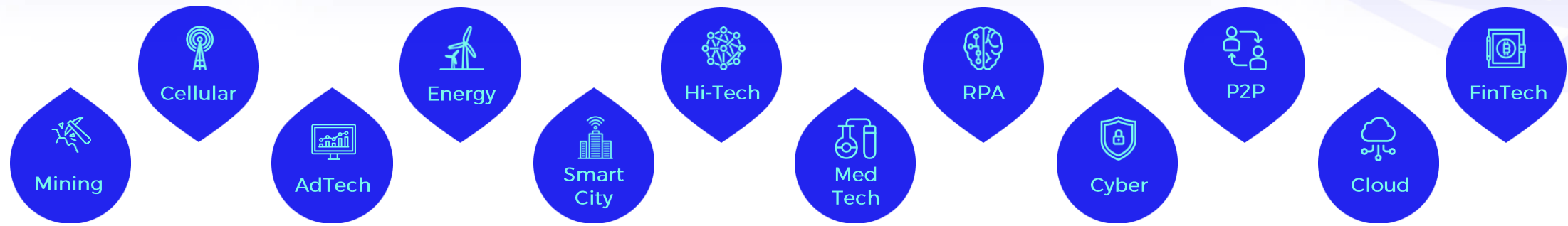
01

Customer Profiling for Cellular Companies – Analysis of CDR

knowledge representation of large-scale networks analytics and graph mining

02

Business process discovery mechanism

Auto-Discovery of business processes that have high probability to be automated

03

Classification and clustering of incoming cybersecurity alerts

Highlighting most relevant alerts and group them accordingly

04

Employees attrition predicative algorithm

Highlight key reasons for possible attrition and preparing mitigation plans

05

Content Tagging for Content Discovery Platform

Automatic discovery and labeling of restrict content

06

Anomaly detection of objects in cameras

Discovery of abnormal behavior of objects in cameras positioned in public zones

07

Optimizing HR Recruitment process of a SaaS-based platform

Development of ML algorithm to predict candidate compatibility

08

Prediction of Facebook Campaigns Performance

ML Algorithm outperforms a human analysts prediction

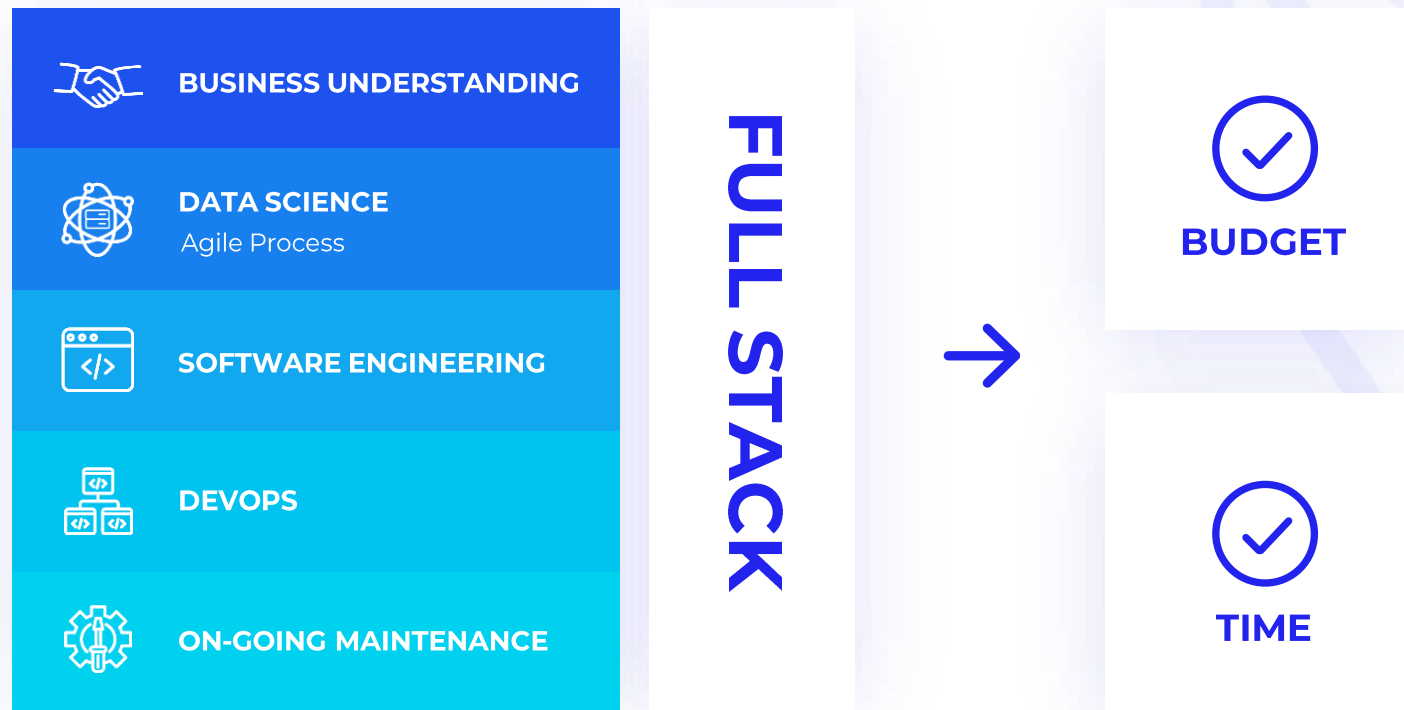**09**

Cloud Customers Segmentation

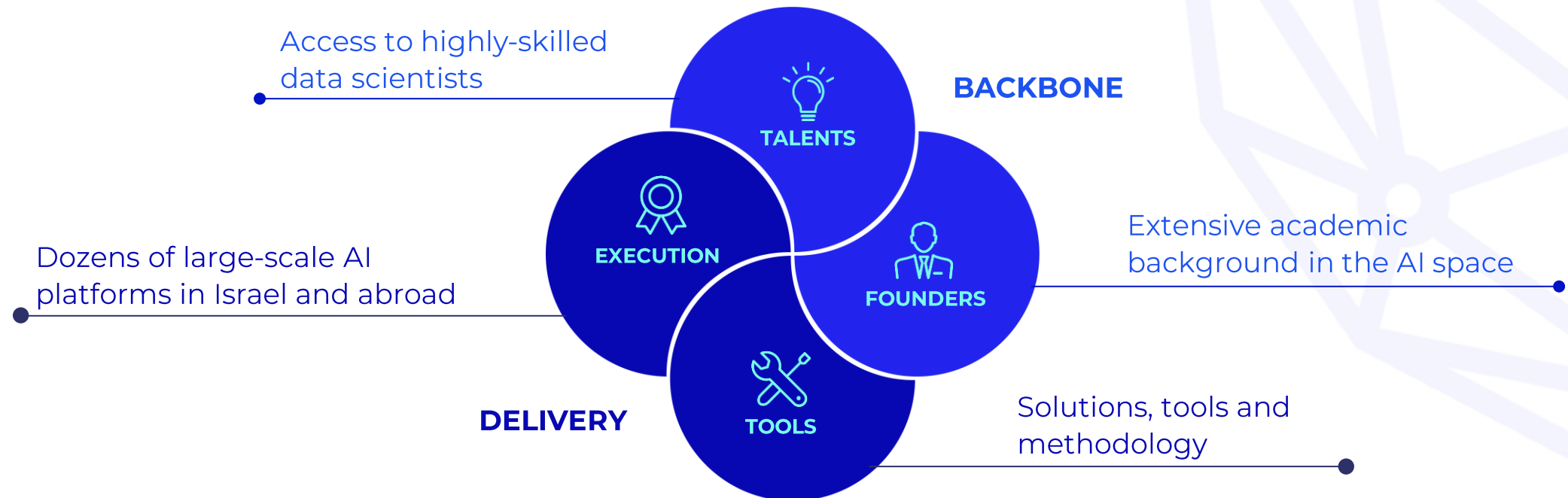A model and decision-support tool of customer profiling and segmentation

10

P2P Credit Scoring

Developing a ML algorithm to validate an existing underwriting process

www.datascience.co.il

DATA SCIENCE GROUP

Full Stack Services

BUSINESS UNDERSTANDING

DATA SCIENCE
Agile Process

SOFTWARE ENGINEERING

DEVOPS

ON-GOING MAINTENANCE

FULL STACK

→

BUDGET

TIME

DATA SCIENCE GROUP

Our Vision – Your AI Partner

Business value = Powerful Tools + Proven Process + Leading AI experts

Access to highly-skilled data scientists

BACKBONE

TALENTS

EXECUTION

Dozens of large-scale AI platforms in Israel and abroad

FOUNDERS

Extensive academic background in the AI space

DELIVERY

TOOLS

Solutions, tools and methodology

www.datascience.co.il

DATA SCIENCE GROUP

The Data Science Journey – Your AI Partner
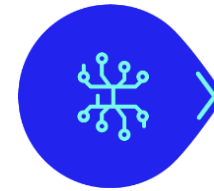
VISION DEVELOPMENT

PLANNING

DATA SOURCES

DATA SCIENCE TEAM

METHODOLOGY & TOOLS

AI/ML APPLICATIONS

DEPLOYMENT

MONITORING

e-volve

01 STRATEGIC → 02 TACTIC → 03 ACTIONABLE → 04 PRODUCTION → 05 RELIABLE

ML 1.0 — ML 2.0

www.datascience.co.il

DATA SCIENCE GROUP

Our Solution

We help AI-driven companies to monitor and to analyze their AI solutions both in development and production environments
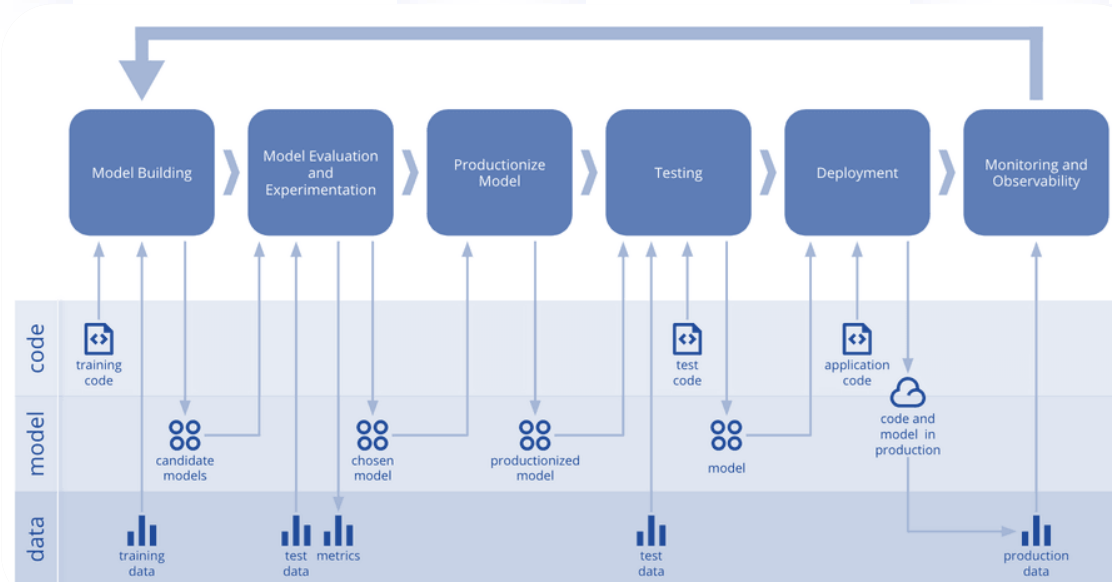
1 DATA SCIENCE VERSIONING + **2** AI MONITORING + **3** MITIGATE = e-volve

DATA SCIENCE GROUP

Data Science Process Analytics

Patent pending platform to perform Data Science

Process Analytics from **Research** to **Production**

to **Monitoring.**

- AI governance

- AI monitoring

- AI auditing

- AI quality assurance

- AI on-going performance management

- AI errors and root cause analysis
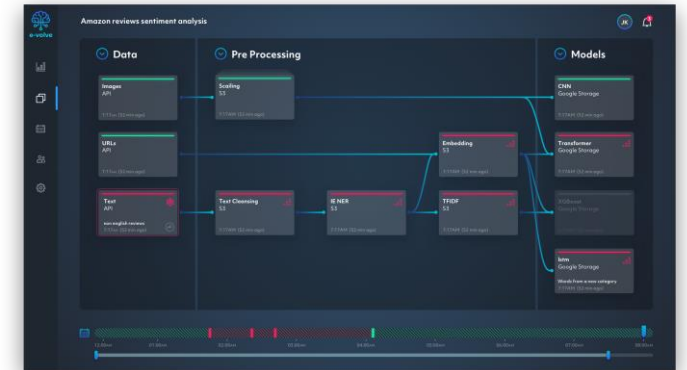
- Decision interpretability/ explainability

e-volve

DATA
SCIENCE
GROUP

Data Science Process Analytics

1 DATA SOURCES → 2 PRE PROCESSING → 3 MODELING TABLE → 4 MACHINE LEARNING MODELS → 5 PREDICTION → 6 MONITORING

R&D. Environment / Production Environment

Monitoring Dataset Shifts

AI-Performance management

AI Quality Assurance

Explainability Interpretability

ORACLE • aws • Google • ▲Azure • Open Source • Customized

API   API   API   **Data Science Pipeline**   API   API   🌳 e-volve

EDDS™ │ Error Driven Data Science

www.datascience.co.il

DATA SCIENCE GROUP

Supplementary Materials

Use Cases Deep Dive

DATA
SCIENCE
GROUP

Content Labeling (Use-Case 1)

Content tagging - content discovery platform

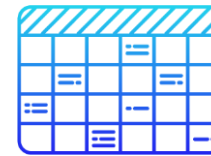Automatic discovery and labeling of restricted content

DATA SET

Over 10M of content items including text and images

TOOLS

Production grade deep learning models implemented in TensorFlow, Image classification, Embedding, Website rating, etc.

MODEL & RESULTS

Novel data embedding and advanced tools implemented in the content review platform

DATA SCIENCE GROUP

Customer Profiling (Use-Case 2)

Customer Profiling for Cellular Companies

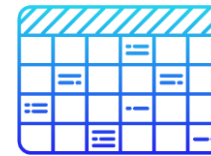Network Analysis of CDRs – Credit Scoring and Marketing Platform

DATA SET

2B records with over 10M nodes and 1.5B edges

TOOLS

Large scale graph database using SNA measures

MODEL & RESULTS

Novel graph mining methodologies.

Classification of default customers and auto-discovery of customer segmentation

Digital Healthcare (Use-Case 3)

Anomaly Detection in EMR – Electronic Medical Record
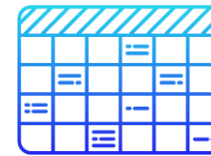Automatic discovery of errors in medical prescriptions

DATA SET
Over eight years of sequential medical prescriptions, admissions, diagnoses and clinical measurements

TOOLS
Data embedding and deep recurrent neural networks

MODEL & RESULTS
Production grade deep learning models .

Real time predictive services of erroneous medical prescriptions for healthcare providers

RPA (Use-Case 4)

RPA – Robotics Process Automation
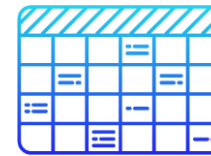Business process auto-discovery mechanism

DATA SET
Over 2M of computer actions recorded from 18 users

TOOLS
Sequences learning methods.

Machine learning algorithms, network analysis tools, and dimensionality reduction techniques

MODEL & RESULTS
Thousands of found sequences have been reduced into few most interesting ones.

Model has been deployed into client's software

DATA SCIENCE GROUP

Human Resources (Use-Case 5)

HR Attrition Prediction
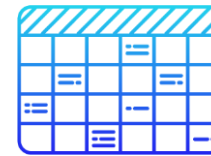Identify employees who are at potential risk of attrition

DATA SET
Over 27K employees - historical data

TOOLS
Supervised learning ensemble model

MODEL & RESULTS
Predictive modeling including web-based application (UI) to help HR managers.

Successfully detected over 90% of the employees that would not leave the company

www.datascience.co.il

DATA SCIENCE GROUP

Peer-To-Peer lending Platform (Use-Case 6)

P2P Credit Scoring
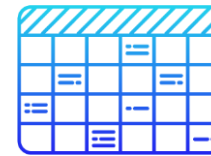
Predict lending requests approval

DATA SET

Using records (4,861 entries) and metrics from weblog files (online service)

TOOLS

Identification of changes the applicant performs on the application form fields

MODEL & RESULTS

Development of ML algorithm to augment the credit risk management (underwriting) process

Cyber Security (Use-Case 7)

Alerts recommendation engine
Classification and clustering of incoming cybersecurity alerts
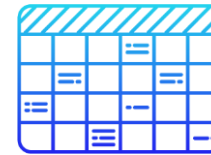
DATA SET
Over 20,000 cyber alerts grouped into 1500 incidents

TOOLS
Machine learning algorithms based on skip-gram based representations (word2vec)

MODEL & RESULTS
Online learning system that follows analyst's preferences, including grouping and auto-discovery mechanisms

DATA SCIENCE GROUP

Process-Based Mining (Use-Case 8)

Optimization system improving production efficiency

Developing an optimization system for the autoclave pressure oxidation circuit
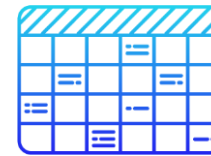
DATA SET
2 years worth of production data

TOOLS
Deep learning temporal architectures implemented in TensorFlow

MODEL & RESULTS
Production grade deep learning model to optimize the autoclave usage as part of a gold extraction process.

DATA SCIENCE GROUP

Clients Segmentation by Cloud Usage (Use-Case 9)

Characterize companies and segment cloud usage

Manage and optimize multi-platform clouds to aid enterprises satisfy their cloud needs intelligently
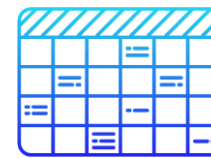
DATA SET

50M entries, 425 features

TOOLS

Spark and H20 for preprocessing and machine learning. Custom software for display, exploration and examination of the results.

MODEL & RESULTS

Customer segmentation provided the basis for a recommendation engine to facilitate efficient cloud usage.

DATA SCIENCE GROUP

Algo-trading Platform (Use-Case 10)

An autonomous end-to-end algo-trading platform

Developed deep learning based platform for future market predictions
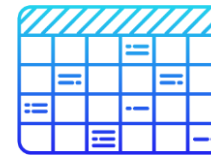
DATA SET

Core fincnacial data

Fundemantal data
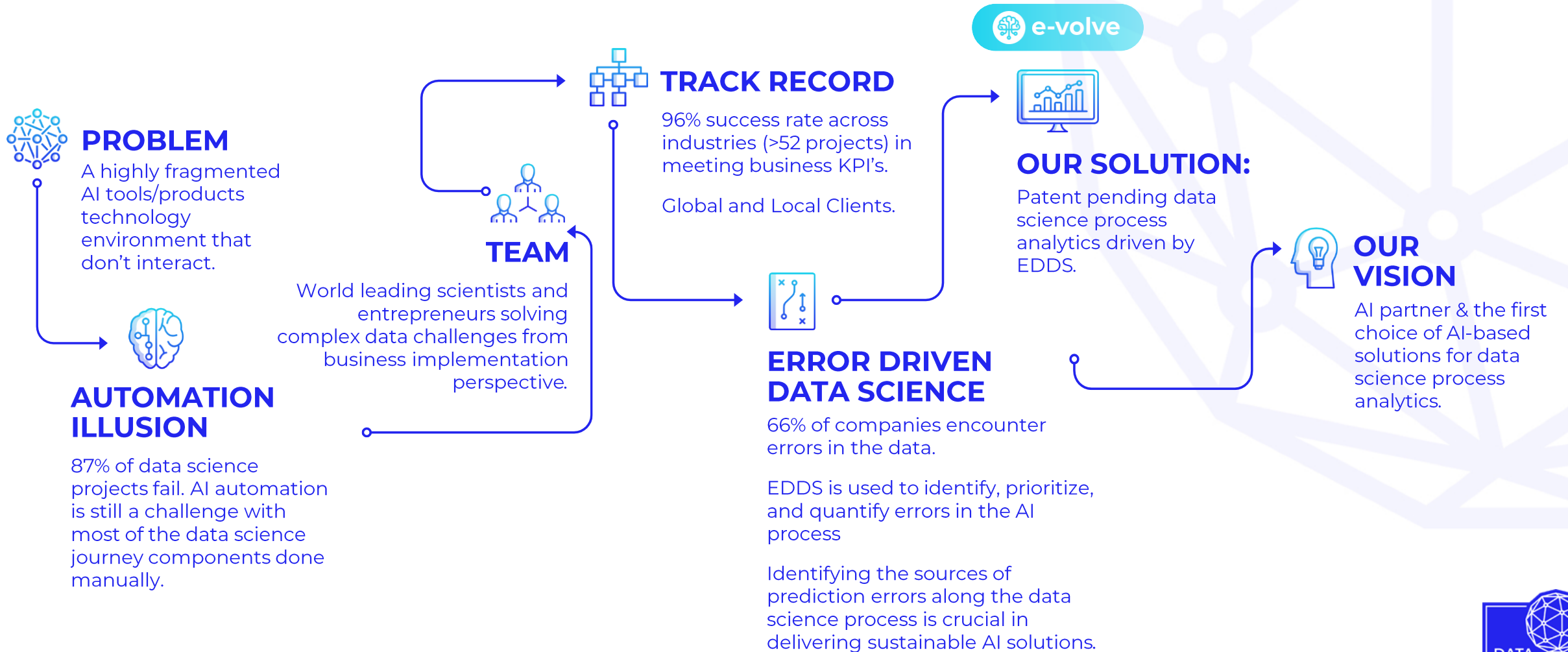
Textual data

Sattelite data

TOOLS

Deep learning temporal architecture

ML optimization algorithm

MODEL & RESULTS

Production grade deep learning system

DATA SCIENCE GROUP

Why DSG?

e-volve

PROBLEM
A highly fragmented AI tools/products technology environment that don't interact.

AUTOMATION ILLUSION
87% of data science projects fail. AI automation is still a challenge with most of the data science journey components done manually.

TEAM
World leading scientists and entrepreneurs solving complex data challenges from business implementation perspective.

TRACK RECORD
96% success rate across industries (>52 projects) in meeting business KPI's.

Global and Local Clients.

ERROR DRIVEN DATA SCIENCE
66% of companies encounter errors in the data.

EDDS is used to identify, prioritize, and quantify errors in the AI process

Identifying the sources of prediction errors along the data science process is crucial in delivering sustainable AI solutions.

OUR SOLUTION:
Patent pending data science process analytics driven by EDDS.

OUR VISION
AI partner & the first choice of AI-based solutions for data science process analytics.

www.datascience.co.il

DATA SCIENCE GROUP

For more information
Please visit our web site
www.datascience.co.il

DATA
SCIENCE
GROUP