H1 2014

Behavior Informatics Laboratories Research and Development Report

Prediction technology based on behavior informatics and application of AI-based technologies to business opportunities in other fields

Presenter: Takeda, Hasuko, Ohnishi

Date: 2014.10.23

http://www.ubic.co.jp

Agenda

Prediction technology based on behavior informatics and application of AI-based technologies to business opportunities in other fields

1. Greetings   *-Masahiro Morimoto*

2. Prediction technology based on behavior informatics and application of AI-based technologies to business opportunities in other fields   *-Hideki Takeda*

3. Prediction technology based on a study of behavior informatics: how it works and what makes it unique   *-Kazumi Hasuko*

4. Meeting the predictive technology needs anticipated by business organizations and looking towards emerging demands of alternative usages of AI-based technology   *-Kenji Ohnishi*

5. Q&A

6. FOC and Behavior Informatics Laboratories office tour

Prediction technology based on behavior informatics and application of AI-based technologies to business opportunities in other fields

Takeda Hideki
Corporate Officer, CTO
Director, Behavior Informatics Laboratories

Prediction technology based on behavior informatics and application of AI-based technologies to business opportunities in other fields

Takeda Hideki
Corporate Officer, CTO
Director, Behavior Informatics Laboratories

Prediction technology based on a study of behavior informatics: how it works and what makes it unique
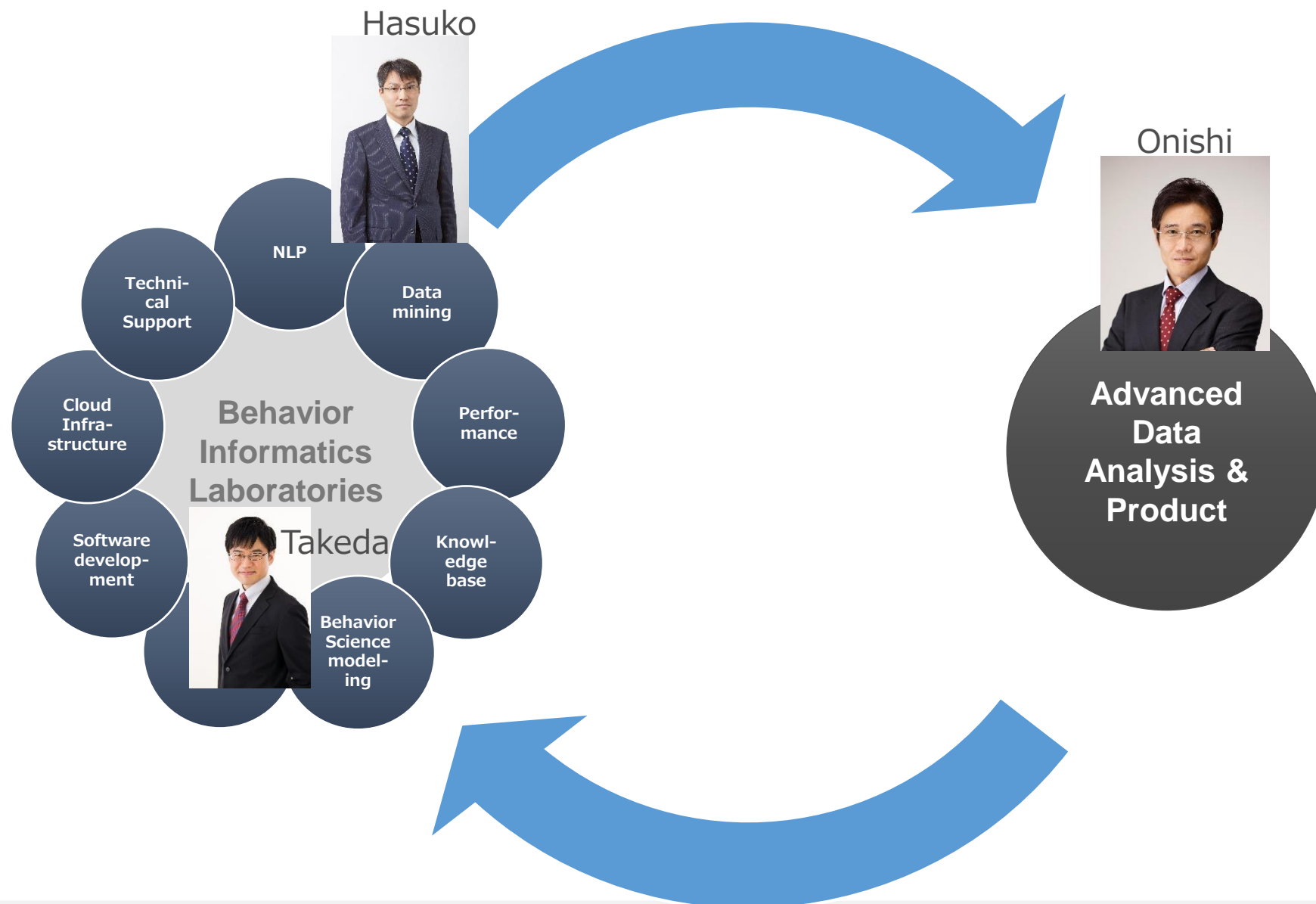
Kazumi Hasuko
Senior Fellow of Behavior Informatics Laboratories, Ph.D. in Science

Meeting the predictive technology needs anticipated by business organizations and looking towards emerging demands of alternative usages of AI-based technology

Kenji Ohnishi
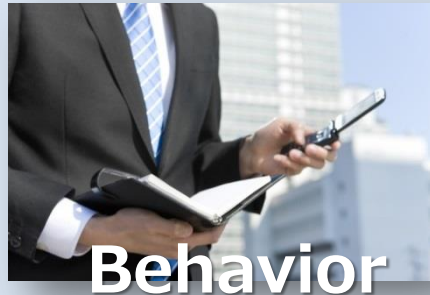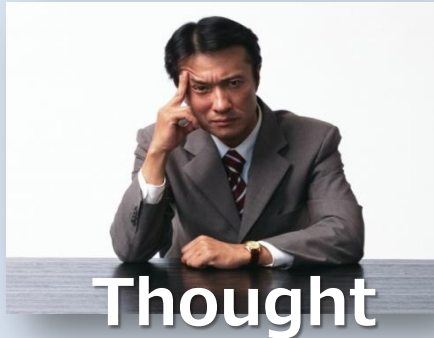Manager, Client Technology Department, Advanced Data Analysis & Product

The workflow cycle of R&D and customer services

Hasuko
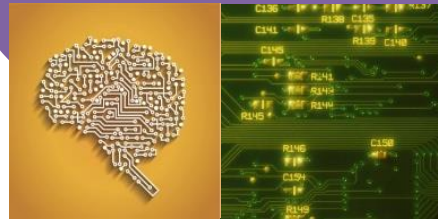
Onishi

NLP

Techni-cal Support

Data mining

Cloud Infra-structure

Behavior Informatics Laboratories

Perfor-mance

Software develop-ment

Takeda

Knowl-edge base

Behavior Science model-ing

Advanced Data Analysis & Product

Concept:
prediction technology based on behavior informatics

Big Data

Thought

Behavior

Analyze →

Behavior Informatics

Behavioral Science

- Psychology
- Criminology
- Sociology

Informatics

- IT
- Statistics

UBIC sees Big Data as an universe of human's thought & behavior

↓

Need to create new analytical method for analyzing human thought and behavior

Concept:
prediction technology based on behavior informatics

Qualitative analysis × Quantitative analysis

By using knowledge of Behavior Science (criminology, social science, psychology), analyze structure of analysis targets from qualitative aspects. Then, evaluate target structure and human behavior by means of statistics. The analysis can be automated.

i.e. criminology

By analyzing structure of corruption (criminology), patterns of human behavior (natural linguistic processing), and communication in connection with misconduct, AI automatically captures indications of criminal.

Analyzing phases: distinguishable action patterns/conversational topics

Phases leading to a cartel	Phases leading to information leakage
Relationship Building	Development
Preparation	Preparation
Execution	Execution

Concept:
prediction technology based on behavior informatics

IT

Technical knowledge of industry

Statistical knowledge

Artificial Intelligence (AI)
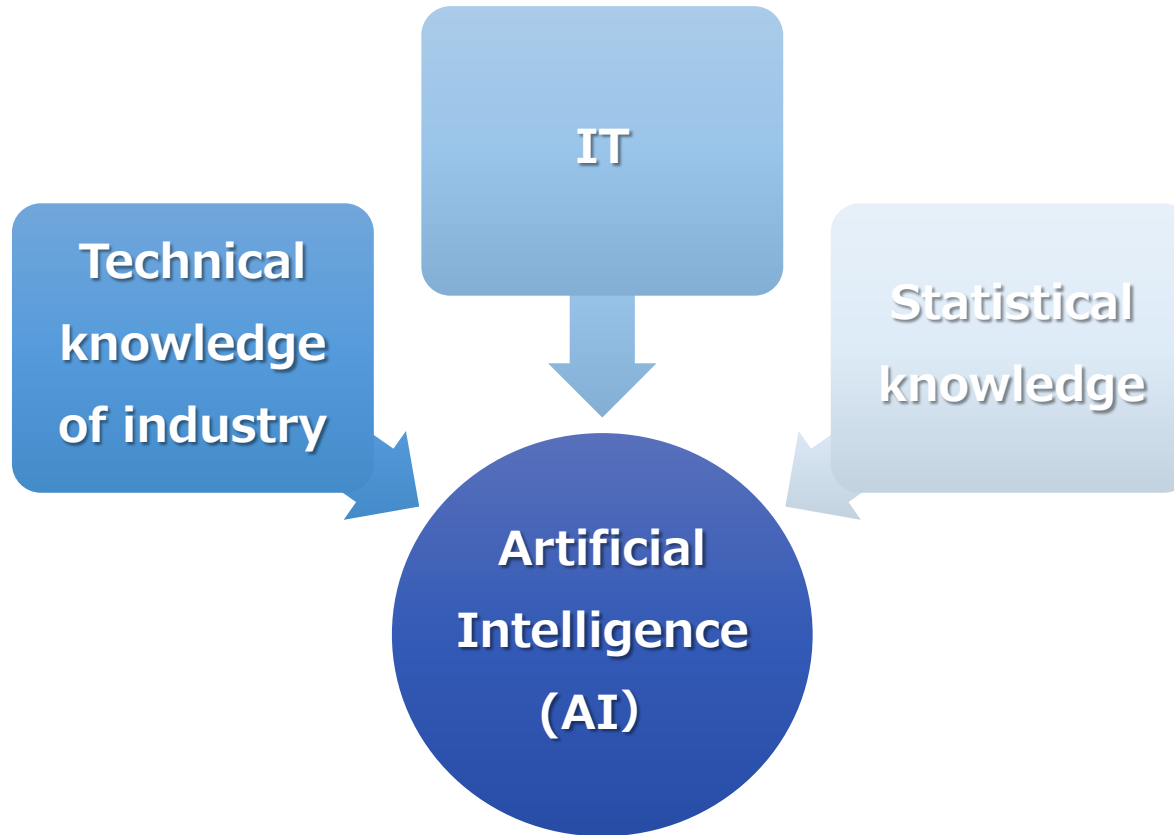
By leveraging knowledge and insights possessed by an expert in a certain business field, the AI can help review a vast amount of data by providing reasonable decisions.

Concept:
prediction technology based on behavior informatics

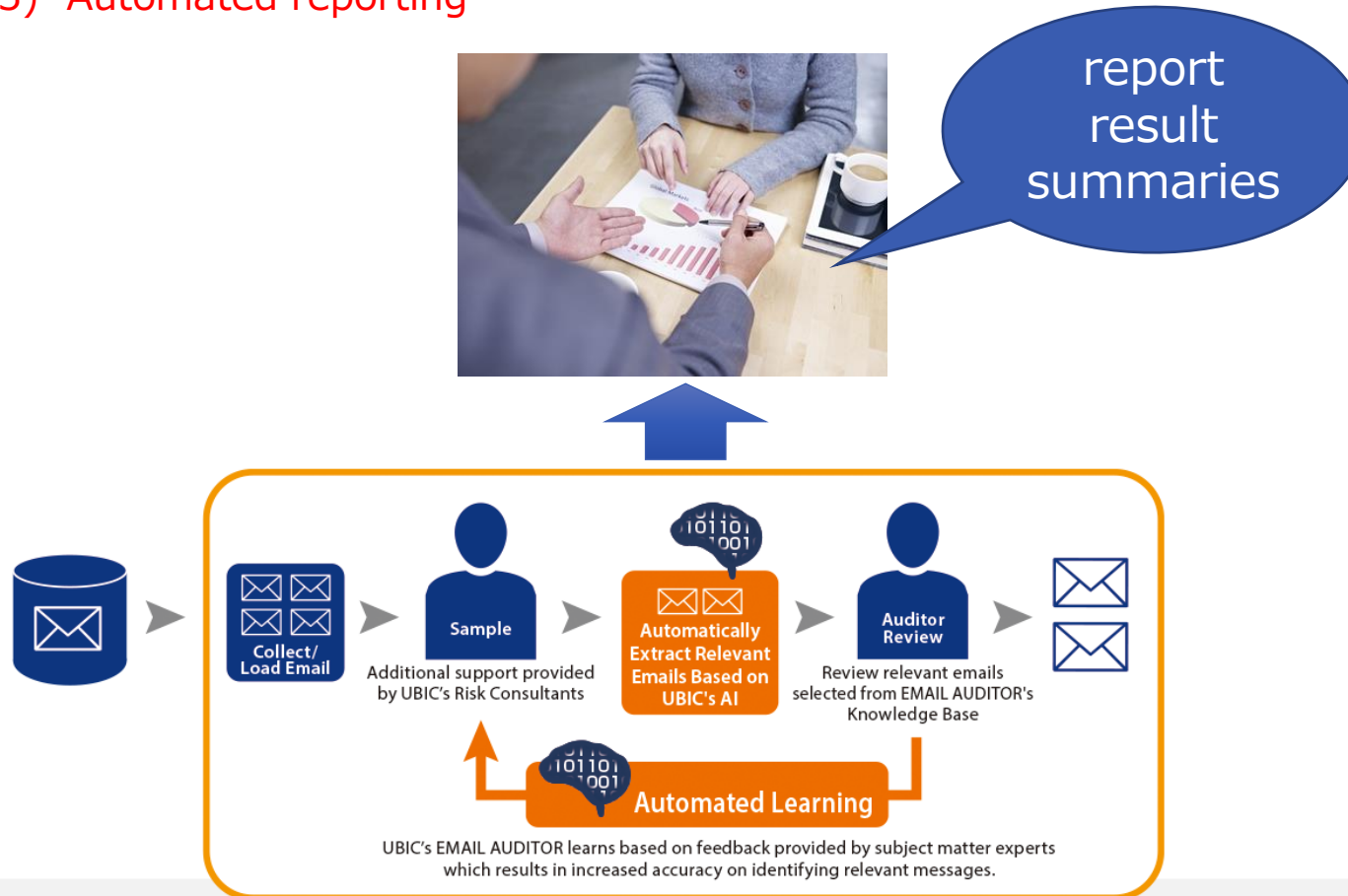AI (VDS) can perform entirely on its own when taking care of business routines.

1) Retrieving emails automatically
2) Automated sorting function

Retrieve messages related to fraud

Conduct an analysis to see which phase the potential fraud falls into

Predict when misconduct is likely going to occur

3) **Automated reporting**

report
result
summaries

Collect/Load Email

Sample
Additional support provided by UBIC's Risk Consultants

Automatically Extract Relevant Emails Based on UBIC's AI

Auditor Review
Review relevant emails selected from EMAIL AUDITOR's Knowledge Base

Automated Learning

UBIC's EMAIL AUDITOR learns based on feedback provided by subject matter experts which results in increased accuracy on identifying relevant messages.
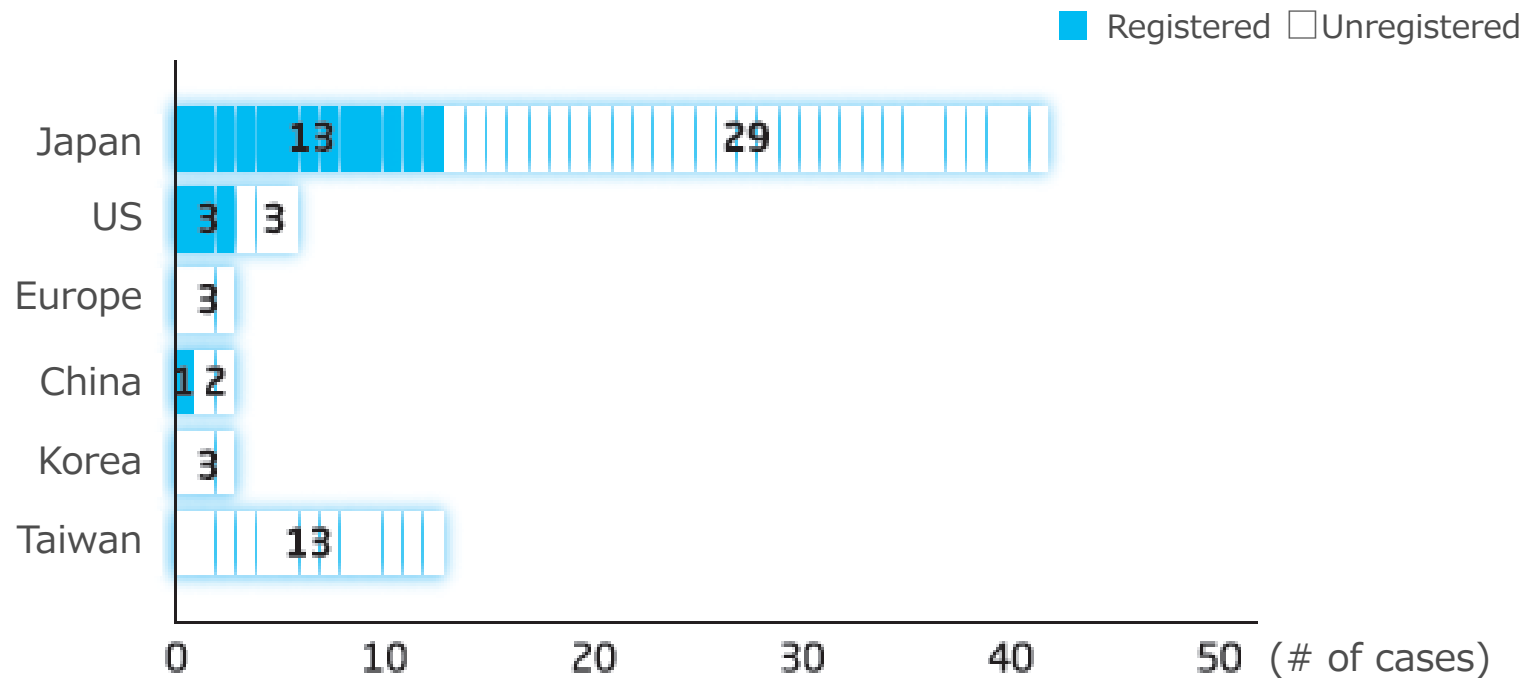
Intellectual Property Strategy
Patent filing status

The patent application/registration status in Japan and abroad, as of August 2014

■ Registered □ Unregistered

Japan	13 — 29
US	3 — 3
Europe	3
China	1 — 2
Korea	3
Taiwan	13

0   10   20   30   40   50   (# of cases)

Intellectual property status
## The new IP strategy
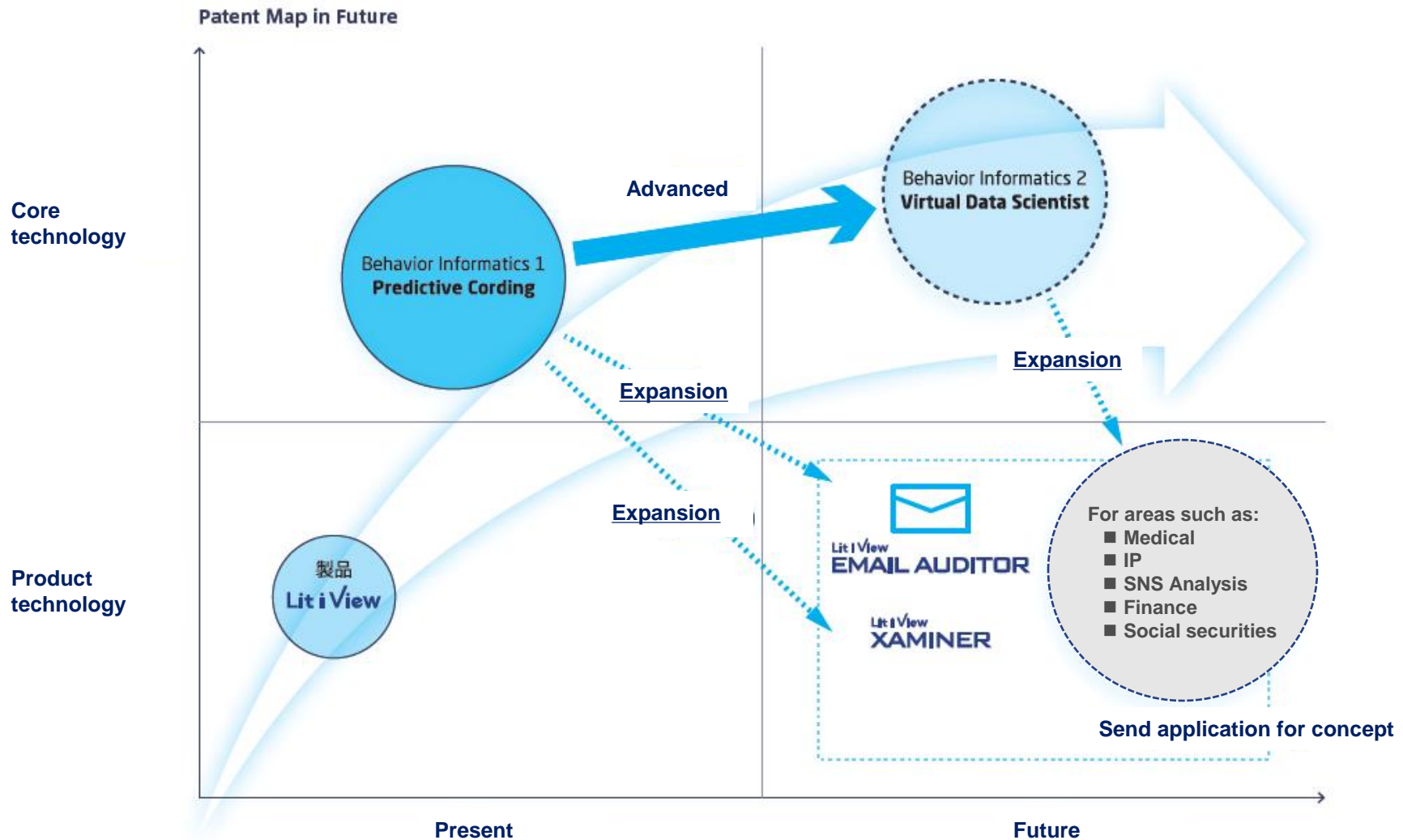
Patent Map in Future

Prediction technology based on a study of behavior informatics: how it works and what makes it unique

Kazumi Hasuko
Senior Fellow of Behavior Informatics Laboratories
Ph.D. in Science

Prediction technology based on a study of behavior informatics: how it works and what makes it unique

○ Risk Prediction
(Prediction technology）

○ Predictive Coding
(AI technology　*Virtual Data Scientist）
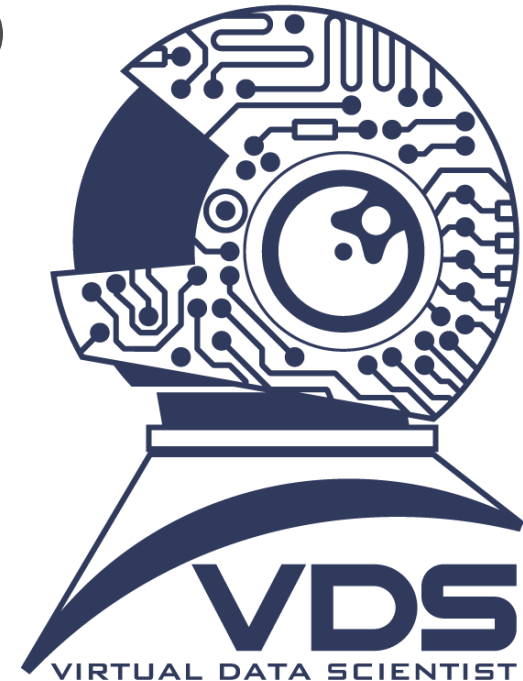
❑ Identify any likelihood of misconduct, before it actually occurs in the near future.

❑ Why can the system predict such risks beforehand?

➢ A database stores knowledge acquired from our careful study on dialogues from past misconducts and fraudulent activities.

➢ The AI teaches itself to accurately read subtle indications of a fraud in a dialogue, by learning and reflecting the knowledge stored in a data base. It will then give its impression on risk factors, such as possibility, situation, and timing.

⇒ Companies can protect themselves from possible misconducts by taking necessary and appropriate measures, beforehand.

The main feature of Risk Prediction: Fraud Phase Analysis

Fraud Phase Analysis allows users to monitor the progress of fraud and analyze data accordingly.
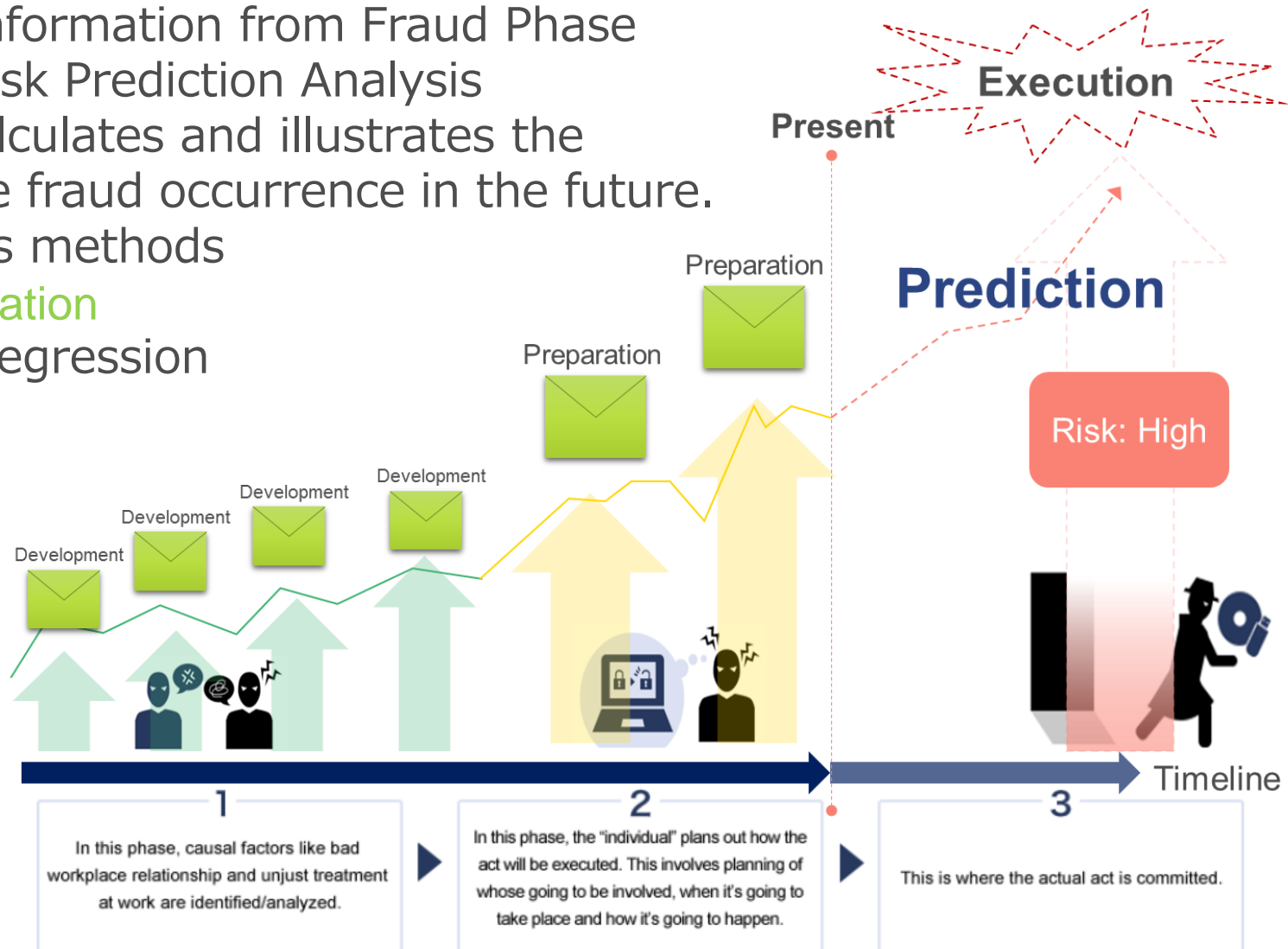

e.g.) Information Leakage

Development

1

In this phase, causal factors like bad workplace relationship and unjust treatment at work are identified/analyzed.

Preparation

2

In this phase, the "individual" plans out how the act will be executed. This involves planning of whose going to be involved, when it's going to take place and how it's going to happen.

Execution

3

This is where the actual act is committed.

The main feature of Risk Prediction:
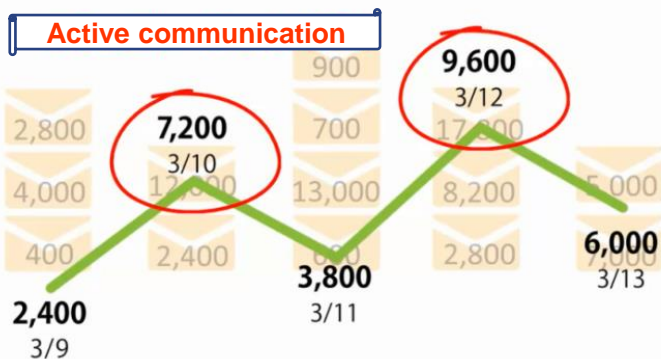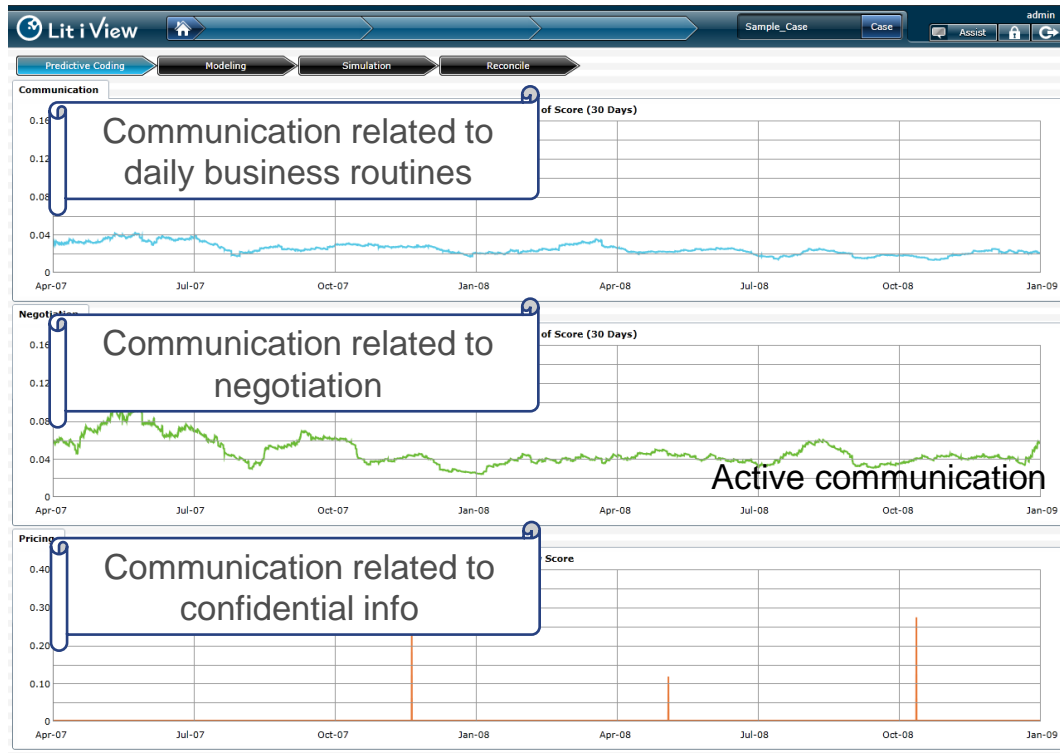# Risk Prediction Analysis

Based on the information from Fraud Phase Analysis, the Risk Prediction Analysis functionality calculates and illustrates the likelihood of the fraud occurrence in the future. Various analysis methods

- Lag correlation
- Logistic regression etc.

Execution

Present

Prediction

Preparation

Preparation

Development

Development

Development

Development

Risk: High

Timeline

1
In this phase, causal factors like bad workplace relationship and unjust treatment at work are identified/analyzed.

2
In this phase, the "individual" plans out how the act will be executed. This involves planning of whose going to be involved, when it's going to take place and how it's going to happen.

3
This is where the actual act is committed.

An example of risk prediction

Email communication are quantified and shown graphically

Communication related to daily business routines

Communication related to negotiation

Active communication

Communication related to confidential info

**Active communication**

Forecast by taking a look at patterns.

Application for Risk Prediction

- Even though all frauds and misconducts differ, there is similarity in their progression from development to emergence.

- A knowledge base is created for each type of past misconducts

  ⇒Risk prediction can be applied to all kinds of misconduct cases.
  i.e. Cartel, FCPA (bribe), information leakage, research misconduct, etc.

Predictive Coding

- ● AI's fundamental technology
  - • Learn expert's knowledge automatically
  - • Analyze documents following the instructions received through supervised learning.

- ● Securing intellectual property rights for our unique technological elements
  - • Mutual Information (patent# 5567049)
  - • Weight Refinement (patent# 5526209)
  - • Progressive Predictive Analyzer
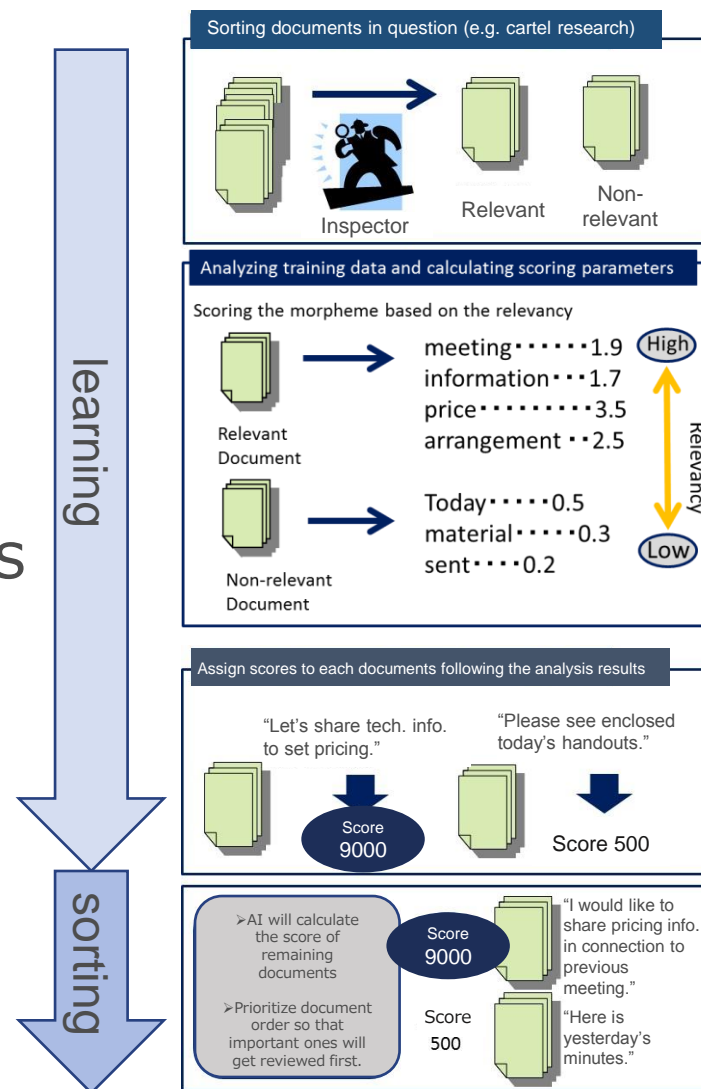
Mutual Information:

- A fundamental information that appropriately expresses the relationship between "significant documents" and "concepts in documents."

- Bayes' rule vs. mutual information

  - **Bayes' rule**
  Search is based on a percentage of significant documents found in a group of messages reflecting concepts.

    ⇒ In other words, this covers only partially.

  - **Mutual information**
  Takes into account more conditions: "messages with or without concept" and "significant or insignificant documents."

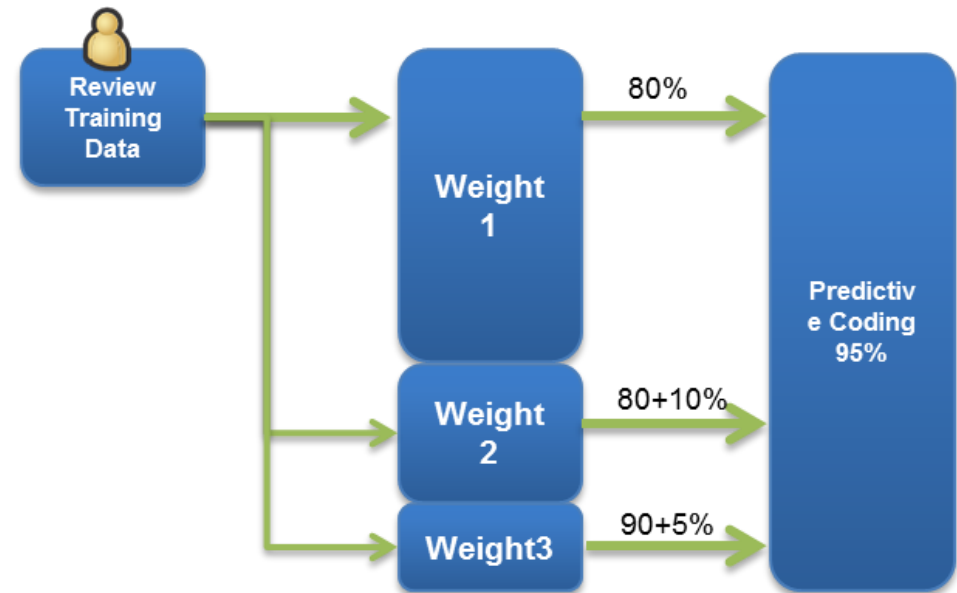   ⇒ Compared with Bayes' Rule, <u>this covers the area more fully.</u>

	Specific concept exists in message	Specific concept does not exist in message.
Document is significant	◯	✕
Document is insignificant	✕	◯

UBIC's unique technological element: Weight Refinement

○ In order to find a broader concept among relevant emails, the weighting is updated automatically to optimize search.

- It will not aim for 100% precision level just by depending on the initially set conditions (Weight 1). If it did, it will eventually lose its precision, instead.

- To include different points of view in a search, the Weights of 2 and 3 can optimize on its own.

⇒ Enhance the ability to extract needed information (Recall rate)

Review Training Data

Weight 1 → 80%

Weight 2 → 80+10%

Weight3 → 90+5%

Predictive Coding 95%

The positioning of UBIC's technology

|	Core technology (elements)	Features/comparison
Virtual Data Scientist (UBIC)	Mutual Information Weight Refinement Progressive Predictive Analyzer	• A comprehensive weighting (enhance recall rate) highly efficient learning capability • Data base are created from reflecting implicit knowledge of expert (training data) so that it is easy to expand them to various applications.
Watson (IBM)	DeepQA	• Processing flow is directed towards a purpose. (Different to computer programs such as neural networks) • It runs for the purpose of "answering questions"
Google	Deep Learning	• Increase layers of neural network. • Processing image and sound • Actualize it by using substantial resources. (Learn from 1,000 PCs over a period of three days)
Pepper (Softbank)	(non-disclosure)	• Perceptive to voice dictation • template/matching (Use fixed phrase when needed)

Meeting the predictive technology needs anticipated by business organizations and looking towards emerging demands of alternative usages of AI-based technology

Kenji Ohnishi
Manager
Client Technology Department
Advanced Data Analysis & Product

On-Site Needs and vision of Predictive Analytics/Artificial Intelligence

○ Various Types of Internal Fraud & its impact
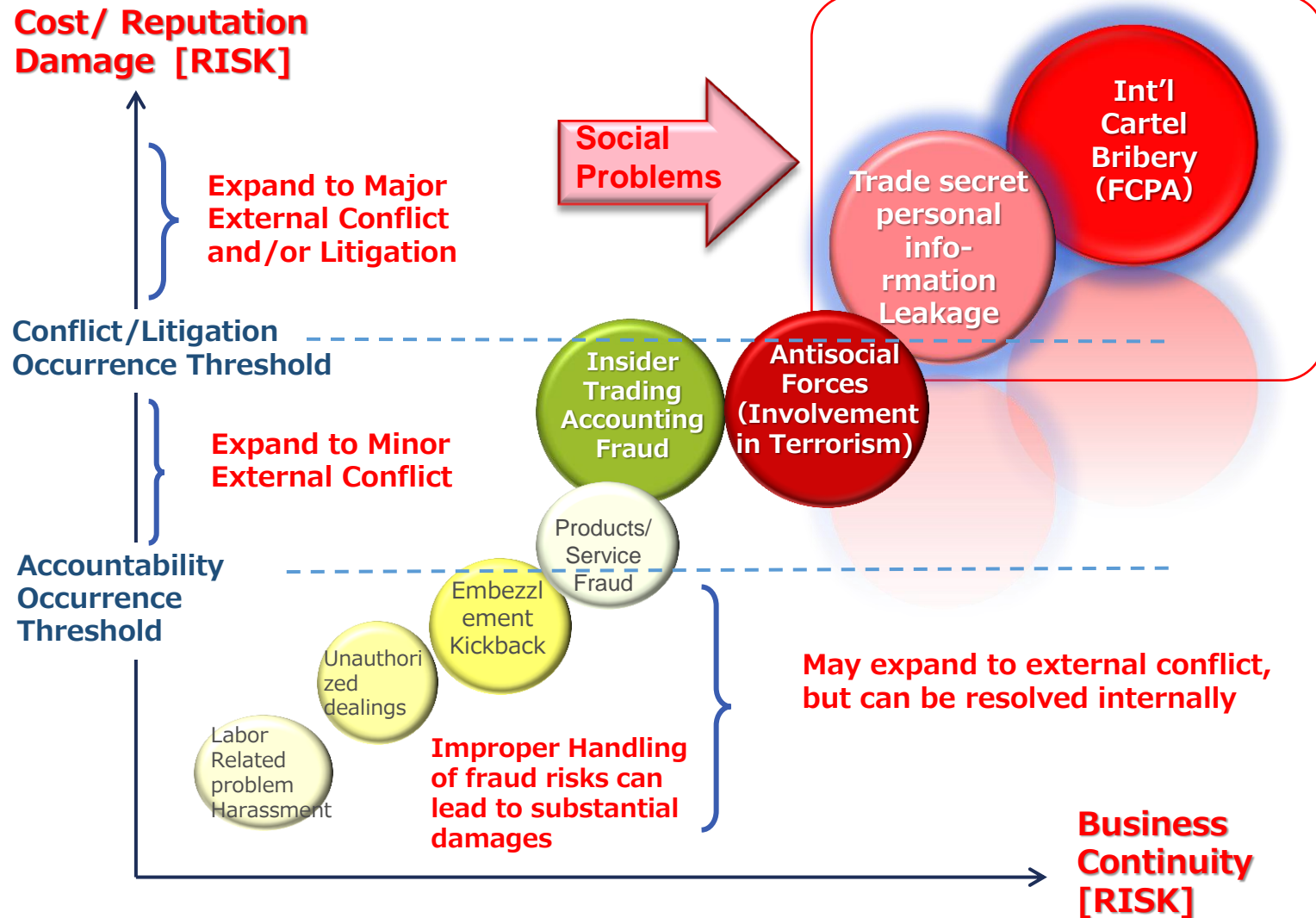
On-Site Needs and vision of Predictive Analytics/Artificial Intelligence

○Difference between Post Incident and Pre Incident

1．Post Incident Handling
・Employee(s) who committed fraud are subject to investigation
・Only PC and File Server are subject to investigation
・Specific illegal issue has already been identified

Target and scope of Investigation can be identified

2．Pre Incident Audit
• Any employee/division of any issues can be subject to investigation(Can't identify Key Words)
• Unspecified employees are subject to investigation(A massive data is subiect to investigation)
• Types of Fraudulence varies along with time (Target issue to be audited
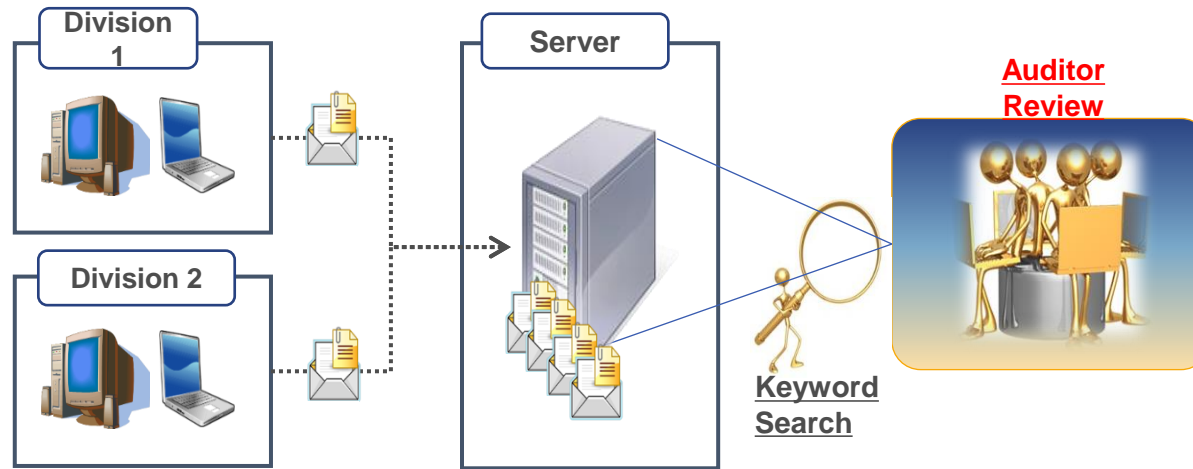

Revolutionary Change in Auditing Methodology Needed to be made!!

On-Site Needs and vision of Predictive Analytics/ Artificial Intelligence

○Issues with Conventional Email Auditing Software

Division 1

Division 2

Server

**Auditor Review**

Keyword Search

Issues:

① A <u>massive search hits</u> can be resulted in as review target
② <u>Critical emails might not be retrieved</u> by too narrow downed KW
③ Need to change Keyword set in accord to the subject matter of the audit
④ Search results are **not prioritized** in order of relevance (**No priority Order**)
⑤ Privacy issues of internal audit

Too much burden for companies and was not effective for "cost-benefit performance."

Solution ⟹ **Prediction Analytics based on Behavior Informatics and AI Technique**

On-Site Needs and vision of Predictive Analytics/ Artificial Intelligence

○ Mechanism of **EMAIL AUDITOR**
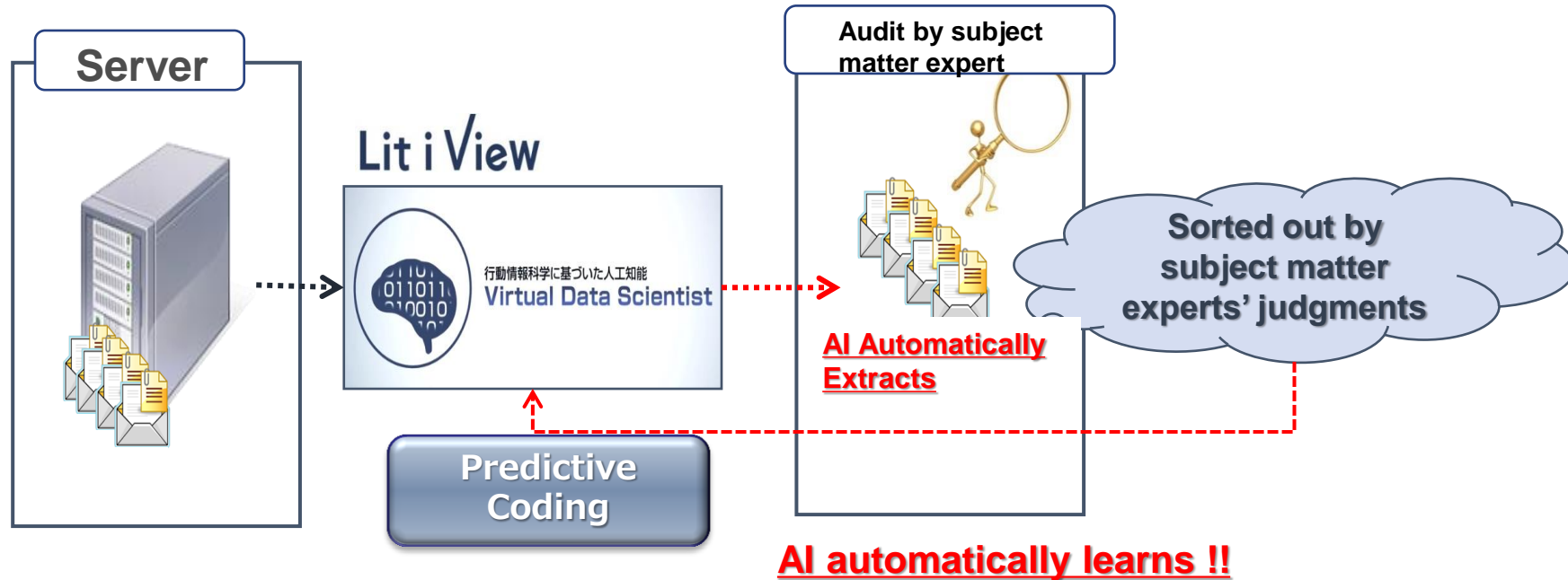
Effects :

・ **AI automatically** detects relevant emails that auditors need with a high degree of accuracy

・ **AI automatically learns** investigation methodology, or implicit knowledge of the subject matter experts, and **improves efficiency**

・ **AI automatically prioritizes** and scores all data **without interrupting any experts'** routine duties

○Risk Prediction: On-Site Needs of Our Clients

・ Demand for EMAIL AUDITOR is increasing amongst our clients as it allows them to detect indications of fraud likely to occur in the near future and allows them to take proper counter measures before it happens.

case 1. To prevent any *Cartel related activities* from
ever happening

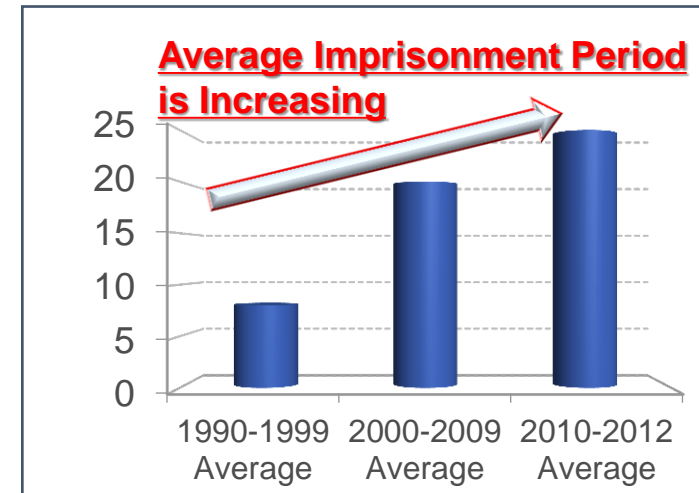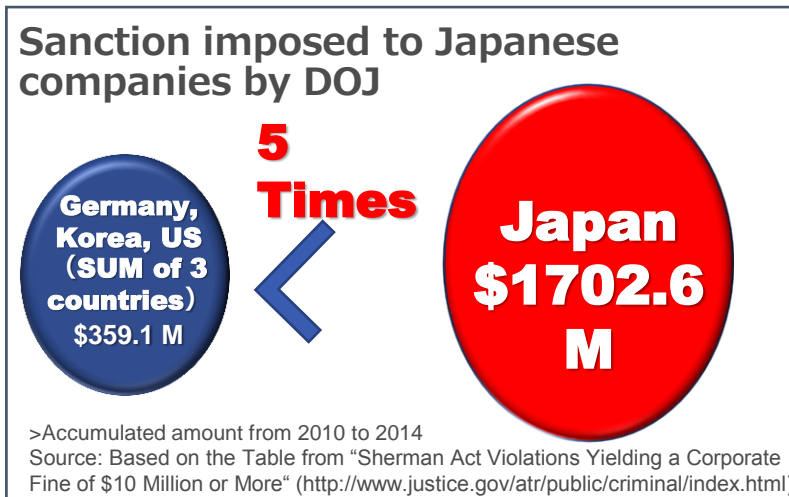case 2. To prevent any Information Leakage from
ever happening

case 3. To prevent employee misconduct by
maintaining employee satisfaction

On-Site Needs and vision of Predictive Analytics/ Artificial Intelligence

○ Risk Prediction: On-Site Needs of Our Clients

case 1. To prevent any *Cartel related activities* from ever happening

DOJ's investigation into potential antitrust violation has expanded even to China and EU today. Once the charge is sentenced, tremendous amount of sanction will be imposed, which impedes business continuity as it requires the business to monitor employees to prevent any engagement of cartel related activities.

Sanction imposed to Japanese companies by DOJ

Germany, Korea, US (SUM of 3 countries) $359.1 M

5 Times

Japan $1702.6 M

>Accumulated amount from 2010 to 2014
Source: Based on the Table from "Sherman Act Violations Yielding a Corporate Fine of $10 Million or More" (http://www.justice.gov/atr/public/criminal/index.html)

Average Imprisonment Period is Increasing

1990-1999 Average	2000-2009 Average	2010-2012 Average

Solution → **Use AI to prevent employees from getting involved in improprietary external communication**

On-Site Needs and vision of Predictive Analytics/ Artificial Intelligence

○ Risk Prediction: On-Site Needs of Our Clients

case 2. To prevent any Information Leakage from ever happening

Unfair Competition Prevention Act was revised and protection of trade secrets was strengthened, but on the other hand, infringement of trade secrets is getting easier for the sake of today's highly developed IT and network. Moreover, once it is infringed, it spreads instantly and might affect irreparable damage to the company to the extent of having no measure to prevent a information leakage.

**Solution** → **Use AI to prevent any Information Leakage from ever happening**

On-Site Needs and vision of Predictive Analytics/ Artificial Intelligence

○ Risk Prediction: On-Site Needs of Our Clients

case 3. To prevent employee misconduct by maintaining employee satisfaction

Labor turnover rate is too high to disregard by considering a loss and drain of human resources

- It takes much more time for a successor to deploy performance equivalent to a predecessor even with handover materials
- Resignation of a well experienced employee discourages its organization as a whole
- Accompanied by a risk to leak a know-how if a resigned employee joins in some other company in the same business

Average loss of new graduates resign at the 3rd year

1500
1000
500
0

1st Yr   2nd Yr   3rd Yr
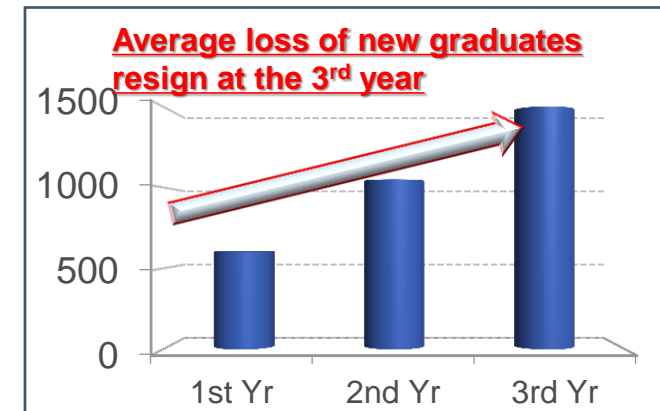
※ Statistics based on our company

Solution → **Use AI to prevent employee misconduct by maintaining employee satisfaction**

Spread technology globally from Japan

From Japan

UBIC, Inc. Proprietary and Confidential.

Creativity, AI

The Artist

Simon Colton is a Professor of Computational Creativity in the Department of Computing of Goldsmiths College, University of London. As an AI researcher who specializes the computational modelling of creative processes, he developed "The Painting Fool", whose art work attracts remarkable attention.

He previously led a research group of the same name at Imperial College, London.

Simon Colton
Professor of Computational Creativity