

第一の夢 AI(人工知能)関連

株式会社 インシリコデータ
湯田 浩太郎

本日のプログラム:

1. 自己紹介: 株式会社 インシリコデータ 湯田浩太郎
2. ◇第一の夢 AI(人工知能)関連
 - * 人工知能の歴史(ルールベース型と機械学習型)
 - * 機械学習型人工知能の現状
 - * 近未来型人工知能の形
3. ◇第二の夢 化合物合成関連
 - * 合成関連支援システムと合成ロボット
 - * 近未来型人工知能の形
4. ◇第三の夢 創薬と医療の統合関連
 - * AIによる、高度かつ迅速、正確な治療への展開
 - * 近未来型創薬と医療の形

日々変化／向上する技術的環境

コンピューター:

ハード; IoT, IT, インターネット, クラウド, スパコン, その他
ソフト; Java, C, HTML, Python, TensorFlow, R, その他

データ解析:

解析手法; より強力で新規のデータ解析手法の展開
データ関連; データの多様化(文字／数値／画像／音声)、
データ量の急激な拡大(データマイニング、ビッグデータ)

人工知能:

ルールベース型人工知能、ニューラルネットワーク、CNN、DNN、
深層学習、強化学習、機械学習、スパースモデリング、他

分析機器:

ハード(種類と機能); クライオ電顕開発、感度／精度向上、他
ソフト; 多次元リアルタイム計測、データ記録様式、他

日々変化／向上する研究手法、ターゲット

創薬関連:

ドラグリポジショニング、マルチターゲット創薬、AI創薬、並列創薬
創薬データサイエンス、テーラーメイドモデリング、他
HTS、ロボット創薬、マイクロ(微量)ドラッグ、iPSを用いた創薬、他

化合物関連:

コンビナトリアルケミストリー、ロボット合成、有機分子触媒、
光／伝導性／分解性ポリマー、有機EL、その他

毒性／環境関連:

計算毒性学、トキシコメタボロミクス、AOP、毒性評価手法、他

バイオ関連:

次世代シーケンサーと解析、様々な遺伝子組み換え実施、
配列解析、発現プロフィール解析、SNP's解析、他

日々変化／拡大する研究手法と情報

創薬

物性

医療

コンピューター

多変量解析
パターン認識

データサイエンス
AI(人工知能)

化合物

物性情報

薬理活性情報

遺伝子情報

化合物情報

毒性情報

文字情報

医療情報

画像情報

数値情報

スペクトル情報

音声情報

1. 人工知能の歴史(ルールベース型と機械学習型)

生命科学分野での人工知能への期待分野と適用／成果

適用分野

化学／創薬

バイオ

医療

人工知能

初期⇒

ルールベース型

現在⇒

機械学習型(深層学習)

期待成果

新薬デザイン、最適化、ドラッグ・リポジショニング、毒性評価、その他

遺伝子解析、遺伝子探索、発現プロファイル解析、SNPs探索、その他

自動診断、画像解析、音声解析、その他

1. 人工知能の歴史(ルールベース型と機械学習型)

時代の変化と人工知能の歴史

□最新/現役/話題の人工知能システム

☆ルールベース型人工知能

DEREK ⇒ 化合物の毒性評価システム

☆コグニティブコンピュータ

Watson (IBM) ⇒ クイズ番組で勝利
医療関連分野で実績を出しつつある

☆機械学習(深層学習)型システム

アルファ碁(グーグル) ⇒


世界のトッププロ棋士(イ・セドル)に4勝一敗で勝利

- ・学習回数: 数千万局 > 三千万局(自己対局)
- ・ルールの自動獲得: 囲碁のルールを自動的に学習した?

時代の変化と人工知能の歴史

☆機械学習（深層学習）型システム


アルファ碁（グーグル）⇒
世界のトッププロ棋士（イ・セドル）に4勝一敗で勝利



勝因？：ルール型対局から一種のパターン認識型対局に変えた
この結果、対局勝利まで今後十年かかるといわれた評価を覆した

□対局の条件と特徴的結果

- ・学習サンプル数：数千万局＞三千万局（自己対局）
- ・ルールの自動獲得：囲碁のルールを自動的に学習した？



上記事実から受ける期待イメージ：

大量のデータを用いて深層学習させると、新規で何らかの
重大なルールが発見でき、新しい研究に繋がるのではないか？

時代の変化と人工知能の歴史

□ 化学分野での人工知能の歴史

当初から現在まで

ルールベース型人工知能

最初に人工知能として開発。
 実用システムや研究システムが多数開発済み。
 技術的にほぼ完成状態。
 長所や欠点が見えている。
 研究的に新規性が殆ど無いため、
論文になりにくい状況

現在注目中

機械学習型
 (ニューラルネットワーク)
 人工知能

機械学習型人工知能

システムとしての実績はないが、今後の展開が期待されている
 ↓
 深層学習
 未知要素が多く、期待値が高
論文になりやすい状況

□ 機械学習型人工知能の全体的歴史と主要トピックス

- 人工知能に注目させたトピックス: **成功事例: アルファ碁**
- 人工知能に注目させたトピックス: **失敗事例: チャットボット「Tay」**
- 人工知能の歴史: **多変量解析／パターン認識と人工知能**

◆**過古**の多変量解析/パターン認識と人工知能との関係

殆ど関連性が無く、全く別の研究分野として扱われてきた
唯一、単層型のパーセプトロンが、分類機として使われていた



◆**現在**における多変量解析/パターン認識と人工知能との関係

現在の深層学習を基本とする人工知能は、ニューラルネットワークが基本。
殆どの場合、ニューラルネットワークは多変量解析／パターン認識に分類される。この結果、両分野の境界は殆ど存在しなくなった。

人工知能の二大手法（ルールベース型と機械学習型）

化合物構造式中心で展開される化学と、
数値／文字中心で展開される自人工知能の
ギャップの解消

□ルールベース型人工知能：過去から現在

化合物の情報学的操作に関する技術は
LHASAシステムの開発過程でほとんど解決済

□機械学習型人工知能：現在

機械学習で化合物を扱う技術は、化学多変量解析／パターン認識
(ケモメトリクス分野)でほとんど解決済

構造式を細かに扱わない深層学習に問題あり

人工知能の二大手法(ルールベース型と機械学習型)

人工知能言語

1958:LISP(List Processor)

1972:Prolog

1994:Python

ルールベース型人工知能

* 第五世代コンピュータ(日本)

**従来型人工知能
実用システム多数**

多変量解析/パターン認識

重回帰、パーセプトロン、PCA、
クラスタリング、他

機械学習発展・新アプローチ

ニューラルネットワーク、
遺伝的アルゴリズム、ファジィ、他

深層学習開発／展開

**新世代
人工知能**

人工知能の変化
時代の変化による

◇ルールベース型人工知能:

人工知能の初期に実施された。

限界; 適用可能な妥当なルールベースを作りこむことが難しい

IoT等の発展により、ビッグデータ時代に入

◇機械学習型人工知能: 最近のメインアプローチ

大量のデータを用いて学習し、自動的に知識表現を獲得

ニューラルネットなどの機械学習手法が発達

最近のディープラーニング等が展開されている

化学分野で現在展開されている人工知能システム

□歴史的に化学関連分野への人工知能適用の歴史は長い

化学分野では数式に乗らない事項が多く、経験則が重要となることが多い

⇒人工知能が活躍する地盤がある

□適用事例は多い

- ・機器スペクトルデータの解析支援システム
- ・有機合成支援システム
- ・毒性予測システム
- ・構造-活性相関支援システム
- ・創薬化学者支援システム
- ・その他

従来より展開されてきた化学分野の人工知能システムは、その展開上化学的なノウハウや考え方等のアナログ的な内容を、デジタルに変換する事が必要

⇒ルールベース型

人工知能による毒性予測関連支援システム： ルールベース型人工知能

DEREK: Deductive Estimation of Risk
from Existing Knowledge

HazardExpert:

RIPT: Rule Induction for Predictive Toxicology

TOX-MATCH:

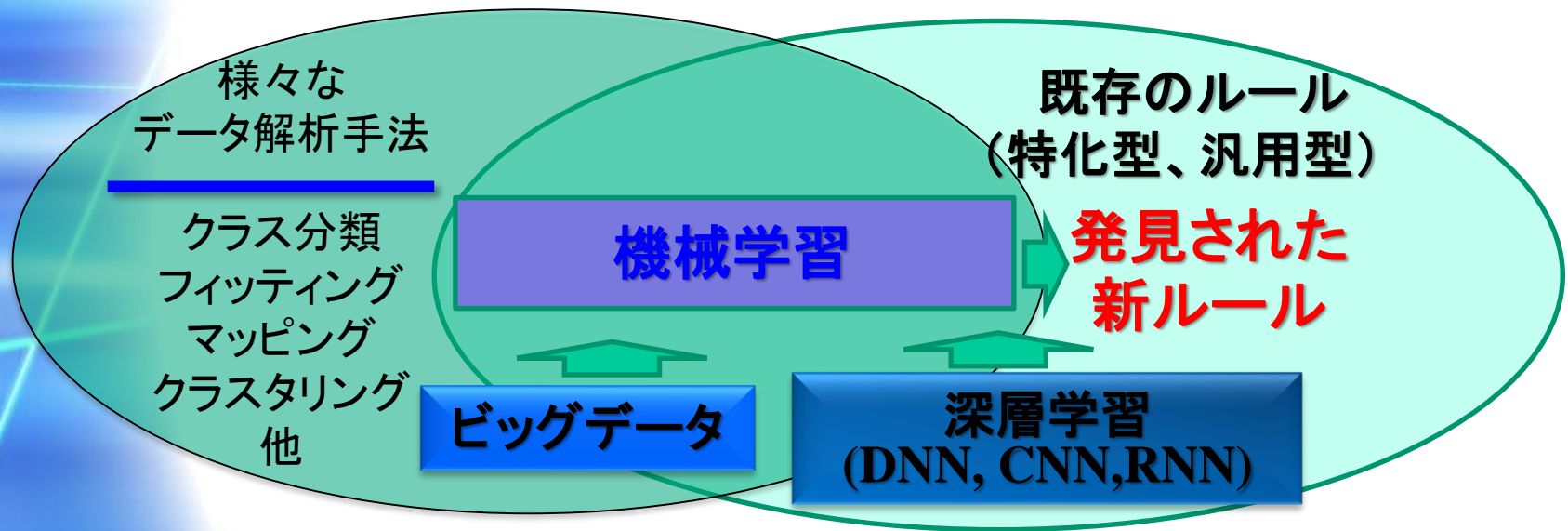
DART: Decision Analysis by Ranking Techniques

現在における多変量解析/パターン認識と人工知能との関係

多変量解析/パターン認識と人工知能は
機械学習により繋がっている

多変量解析/パターン認識

人工知能



1. 専門研究AIから統合研究AIロボットへの展開

例:

- ① 専門機能AIロボット(画像診断)⇒診断精度向上、高速化
- ② 専門機能AIロボット(治療／投薬)⇒治療の実施
- ③ 専門機能AIロボット(デザイン)⇒化合物のデザイン



「統合研究支援AIロボット」

統合AIシステム:(例)

「乳がん」の検診をして、治療計画を立て、投薬し、新規化合物のデザインまで実施するAIロボット

創薬

物性

医療

コンピューター

化学多変量解析 / データサイエンス
パターン認識 AI(人工知能)

化合物

物性情報

薬理活性情報

遺伝子情報

化合物情報

毒性情報

文字情報

医療情報

画像情報

数値情報

スペクトル情報

音声情報

□データやシステムの階層構造

化合物

創薬 物性 医療

種々データ／ビッグデータ

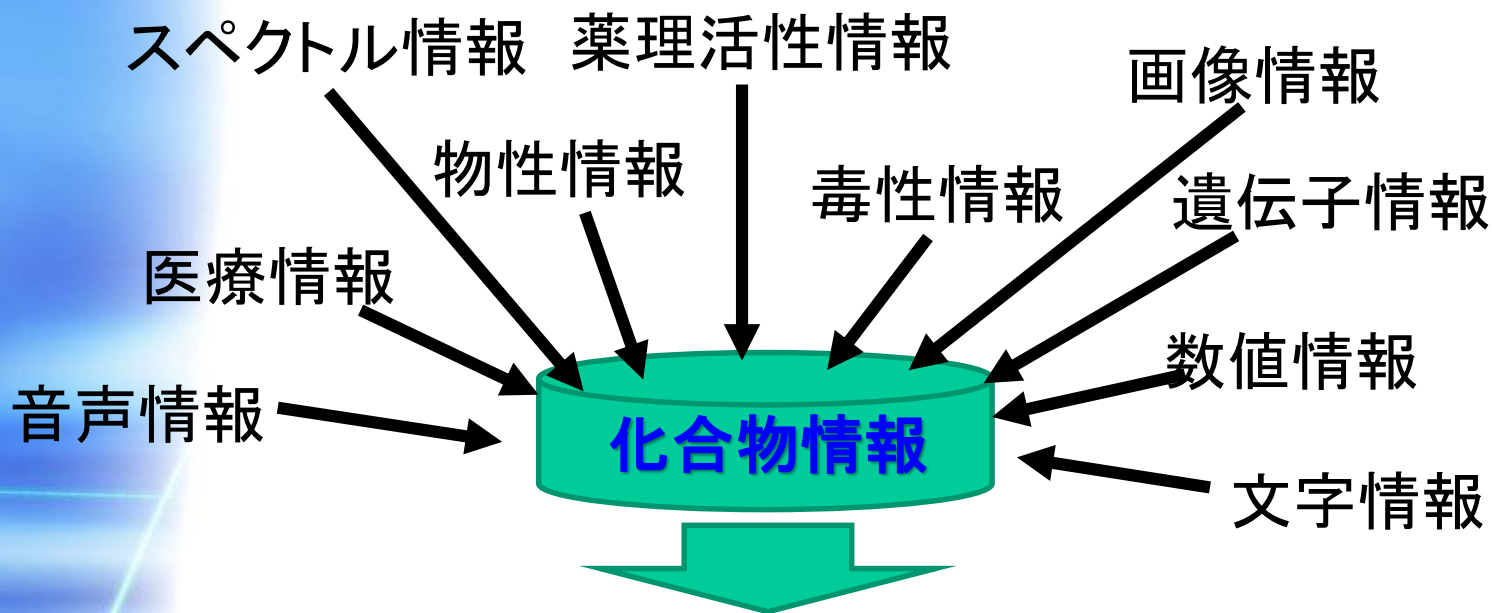
化学データサイエンス

化学多変量解析
パターン認識

AI(人工知能)

コンピューター

□多様な情報、システム、解析ソフト、DB等



化学データサイエンス

化学多変量解析
パターン認識

AI(人工知能)

現在のAIシステムの特徴と欠点：

◇ 単機能システムである

システム内で機能を完結し、最高のパフォーマンスを実現

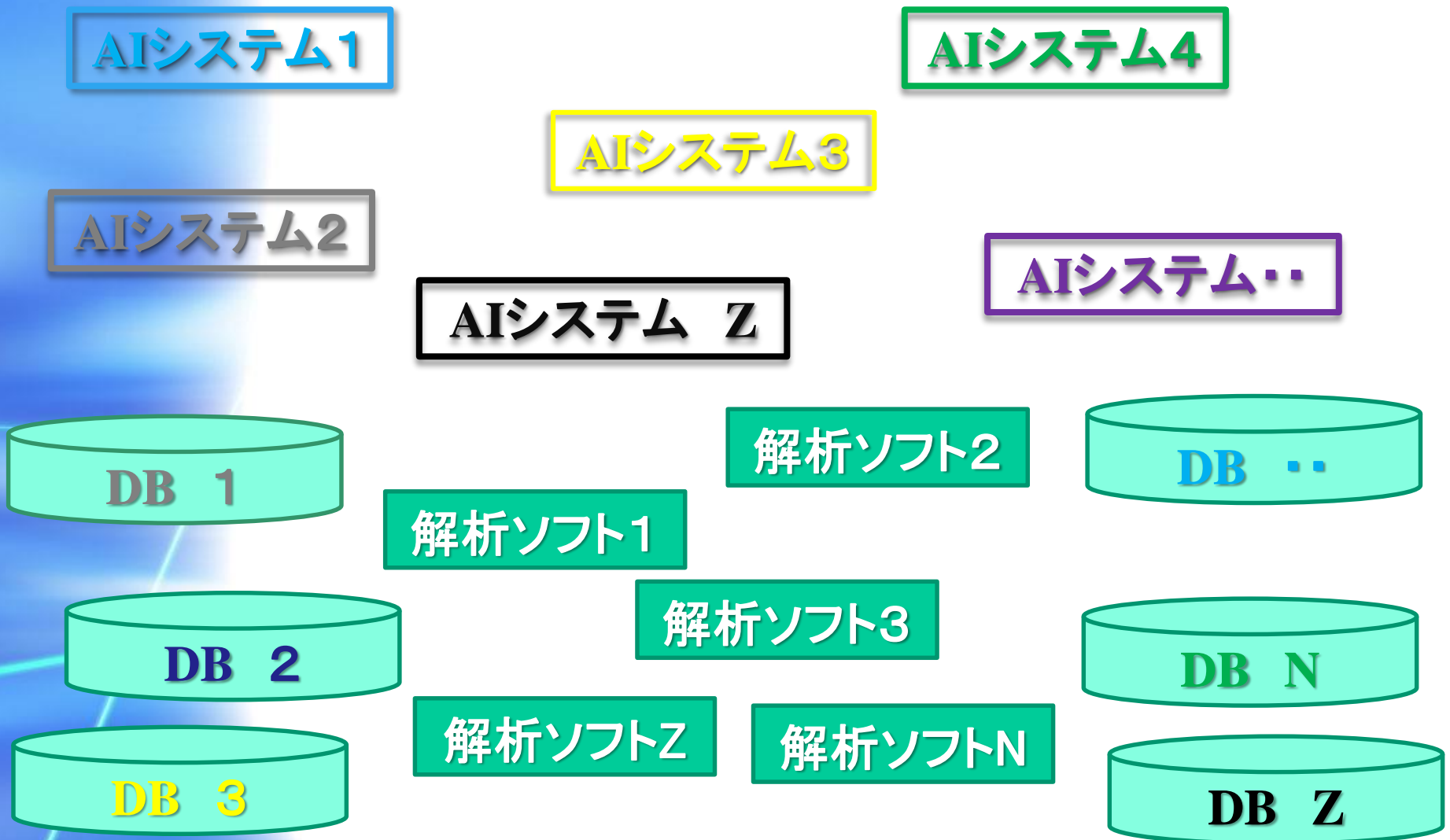
◇ 複数システムとの連携困難

他のシステムとの連携が困難で、システム内での完結を目指す

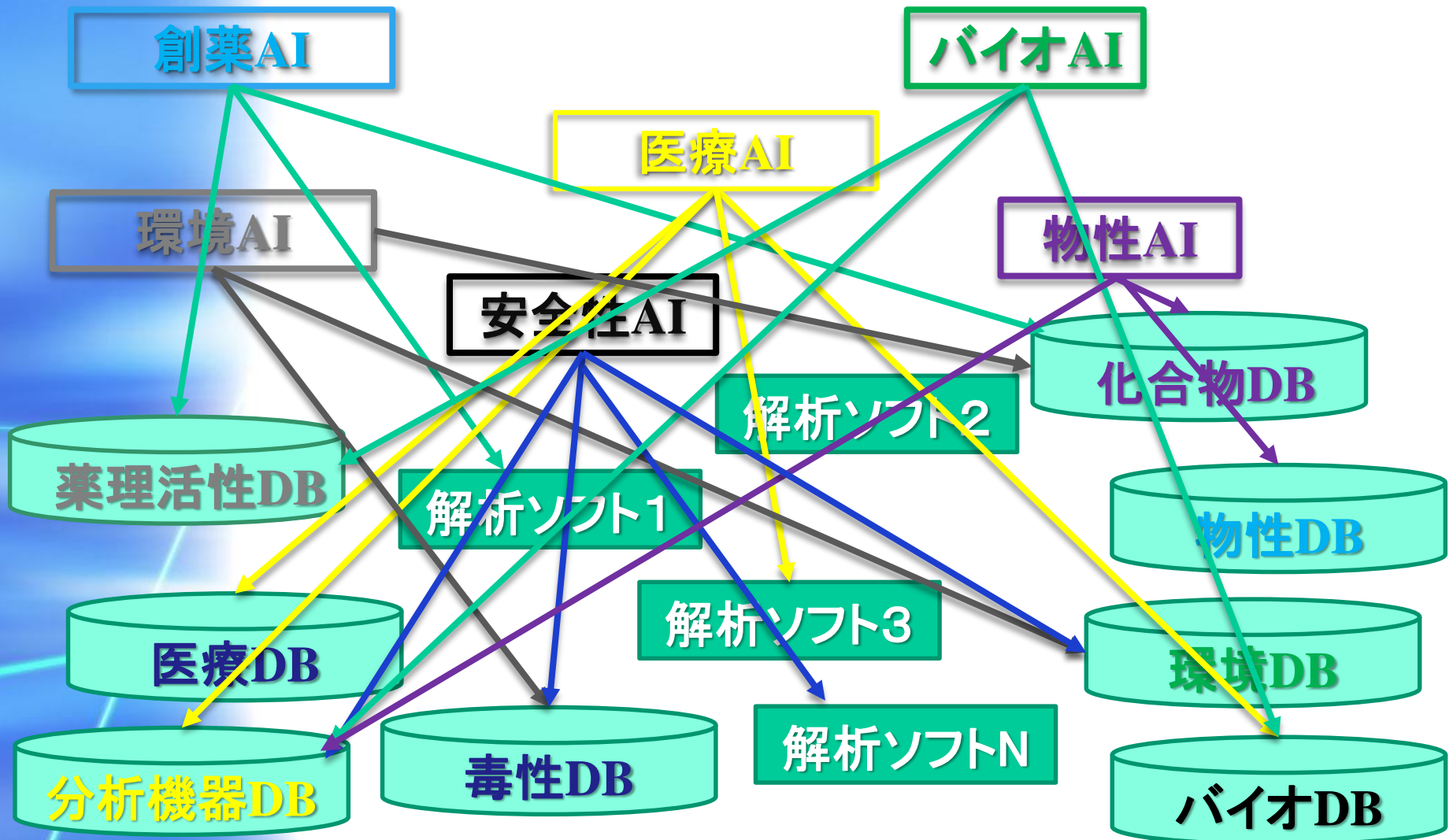
◇ 環境(コンピューターや研究手法)の変化に弱い

独立システムであり、環境の変化に伴う機能追加等が困難

□ 多様なAIシステム、解析ソフト、データベース等



□多様なAIシステム、解析ソフト、データベース等



□多様なAIシステム、解析ソフト、データベース等

長所:

1. 個々のシステムは最高のパフォーマンスを出しやすい
2. システムサイズは小さいことが多く開発しやすい

欠点、限界:

1. より高度／複雑な研究や機能を求めることは困難
2. システム間の連携が困難
3. 新機能システムは新規作成が必要(再利用は困難)
4. 研究等の変化に対応する事は極めて困難(使い捨てとなる)

□未来化学データサイエンスシステム

総合研究実施
AIロボット

将来

より高度なAI研究
臨機応変にAI研究適用
複数AI研究のまとめ

現状

- ・多種多様のAIシステム
- ・多種多様のデータサイエンスシステム
- ・多種多様のビッグデータベース

□現状のシステム環境

AIシステム1

AIシステム3

AIシステム4

AIシステム2

AIシステム..

AIシステム Z

総合研究実施
AIロボット

DB 1

解析ソフト1

解析ソフト2

DB ..

DB 2

解析ソフト3

解析ソフトN

DB N

DB 3

解析ソフトZ

DB Z

近未来におけるAIの形

多種多様な解析テーマに、
配下のAI、解析ソフト、
データベースを能動的に適用

解析
テーマ1

解析
テーマ2

解析
テーマ・

解析
テーマZ

総合研究実施
AIロボット

AIシステム

解析ソフト

DB

□ 専門AIを駆使して大きなターゲット

新テーマ 高度なテーマ 複雑なテーマ



近未来

総合AIセンター

専門
AIシステム1

専門
AIシステム2

専門
AIシステム...

専門
AIシステム...

専門
AIシステムN

専門
AIシステムN+1

ビッグデータ

データサイエンス

現状

化合物、蛋白、遺伝子; 薬理活性、毒性、ADME、物性

ご清聴ありがとうございました

株式会社 インシリコデータ
湯田 浩太郎