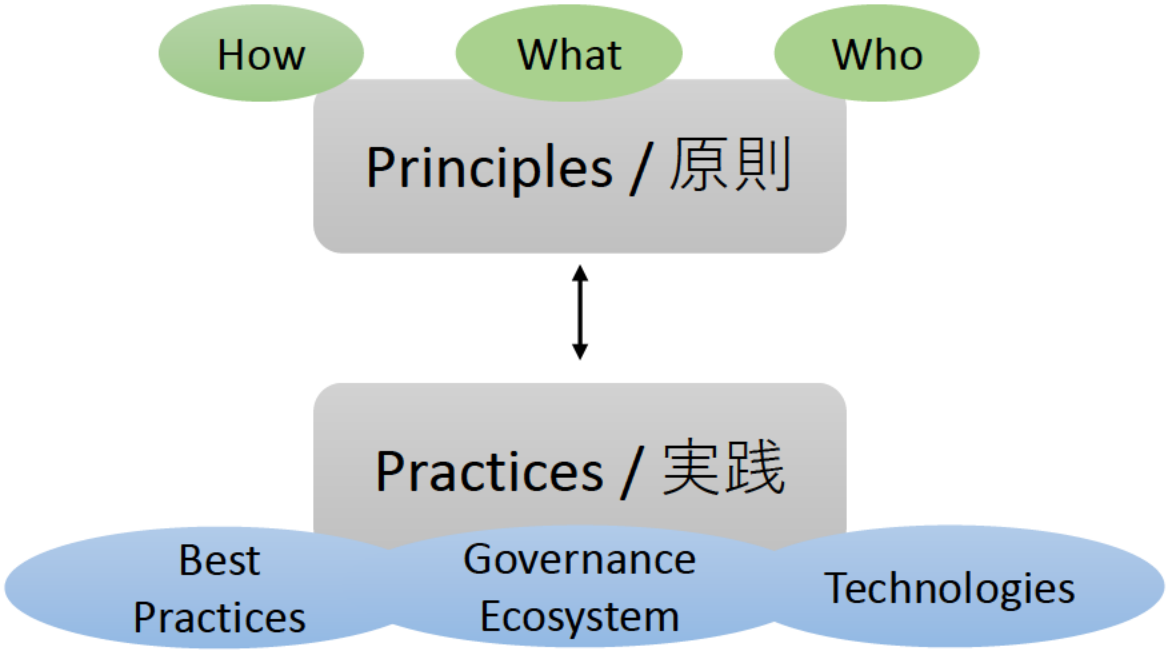
資料１

AI開発原則と実装

AI Principles and Practices

東京大学未来ビジョン研究センター特任講師
理化学研究所AIPセンター客員研究員
江間　有沙


How
What
Who
Principles / 原則
Practices / 実践
Best Practices
Governance Ecosystem
Technologies

What

- Fairness
- Accountability
- Transparency
- Trust
- Explainability
- Sustainability
- Safety & Security
- Privacy
- etc

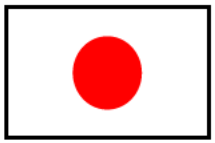
How

- Multi-Stakeholder
- Diversity & Inclusion
- International / Regional

Who

- Academic communities
- Industry Associations
- NPOs/NGOs
- Nation / Region
- International organizations

M: 56.9 %
F : 42.1 %



M: 86.2 %
F : 13.8 %

Male-female ratio of AI principle committee members

Principles / 原則

Focus on areas
- Medicine/ health
- Finance

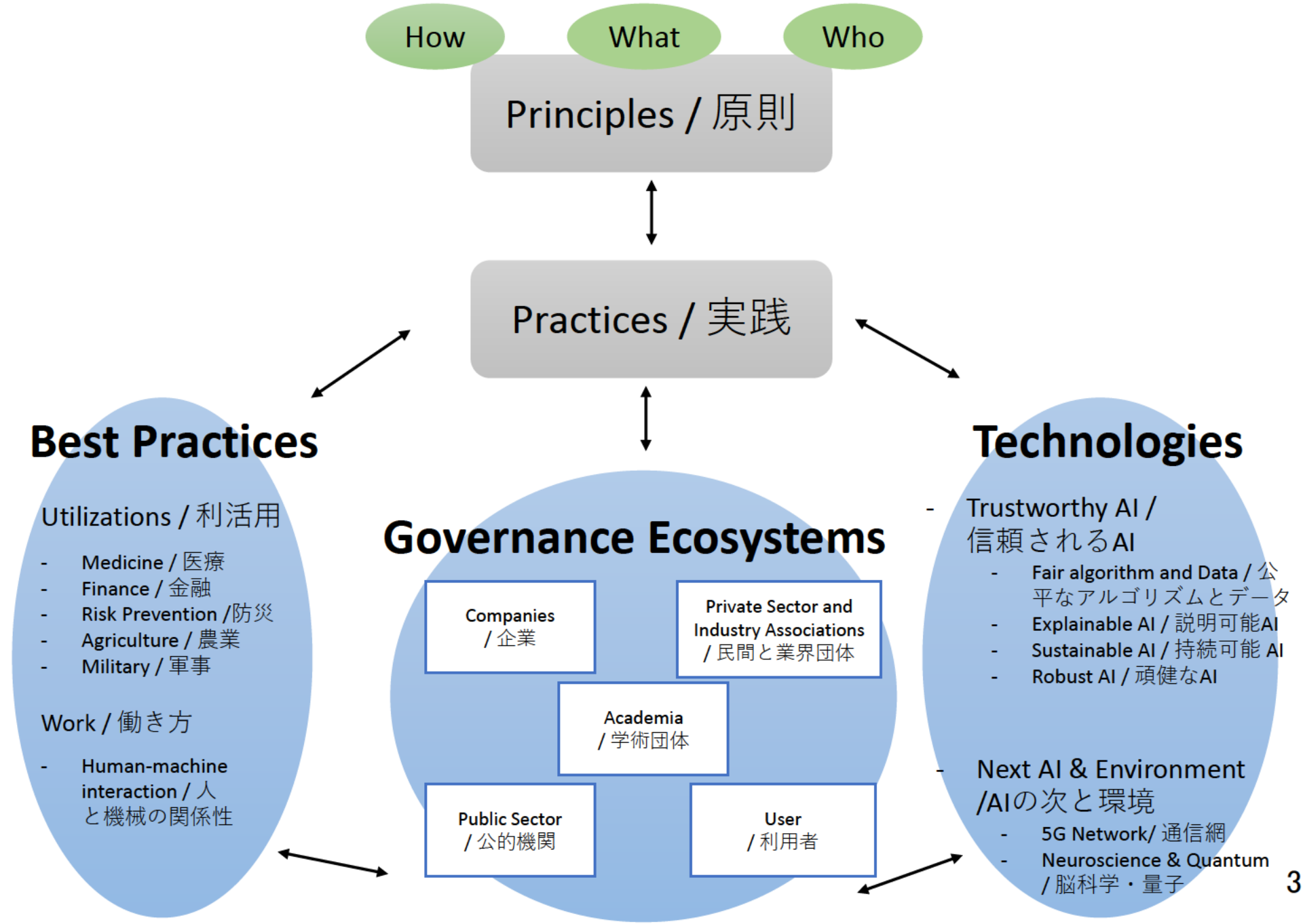
Focus on regions
- Global South

Practices / 実践

Best Practices

Governance Ecosystem

Technologies


How
What
Who
Principles / 原則
Practices / 実践
Best Practices
Utilizations / 利活用
- Medicine / 医療
- Finance / 金融
- Risk Prevention /防災
- Agriculture / 農業
- Military / 軍事
Work / 働き方
- Human-machine interaction / 人と機械の関係性
Governance Ecosystems
Companies / 企業
Private Sector and Industry Associations / 民間と業界団体
Academia / 学術団体
Public Sector / 公的機関
User / 利用者
Technologies
- Trustworthy AI / 信頼されるAI
- Fair algorithm and Data / 公平なアルゴリズムとデータ
- Explainable AI / 説明可能AI
- Sustainable AI / 持続可能 AI
- Robust AI / 頑健なAI
- Next AI & Environment /AIの次と環境
- 5G Network/ 通信網
- Neuroscience & Quantum / 脳科学・量子

Ex. 1) Ecosystem of AI governance /
例1) AIガバナンスのエコシステム

Companies / 企業

- Corporate vision / ビジョン
- AI Principles / AI原則
- Risk Assessment / リスク評価
- Risk Control / リスクコントロール
- Employee/ Employer Education / 教育

Private Sector and Industry Association / 民間と業界団体

- Guidelines / ガイドライン
- Standards / 標準
- Audit / 監査
- Insurance / 保険
- Fact Check / ファクトチェック

Academia / 学術団体

Public Sector / 公的機関

- Hard Law, Soft Law / 法や規制
- Third-Party Incident Committee / 事故調査制度
- Whistleblower System / 内部告発制度

Users / 利用者

- Education / 教育
  - Engineer / エンジニア
  - Experts / 専門家
  - Policy makers / 政策関係者
  - General publics / 一般

研究会：AIガバナンスとその評価

座長	江間有沙（東京大学 未来ビジョン研究センター 特任講師）
検討課題	多様なアクターによる管理・評価の体制の在り方を「ガバナンス」と定義し、どのようなガバナンスの形がありうるのか調査し、信頼されるAIの構築の一助とする。

第４回　AIサービスに係る保険

開催日時	2020年10月26日（月）
内容	話題提供/ディスカッション
テーマ	AIを含むサービスに係る保険
話題提供者	風間啓（損害保険ジャパン株式会社） 永野智也（東京海上日動火災保険株式会社）

第３回　AIシステムにおける監査・保証

開催日時	2020年9月25日（金）
内容	話題提供/ディスカッション
テーマ	AIシステムにおける監査・保証
話題提供者	阿子島隆（Japan Digital Design株式会社） 長谷友春（有限責任監査法人トーマツ）

配布資料はこちら（JDLA会員限定）

第２回　AI倫理ガイドライン

開催日時	2020年8月25日（火）
内容	話題提供/ディスカッション
テーマ	AI倫理ガイドライン
話題提供者	松本敬史（有限責任監査法人トーマツ/研究会副座長）

配布資料はこちら（JDLA会員限定）

開催レポート

第１回　AIガバナンスをめぐる国内外の動向

開催日時	2020年7月31日（月）
内容	本研究会について/話題提供/ディスカッション
テーマ	AIガバナンスをめぐる国内外の動向
話題提供者	江間有沙（東京大学 未来ビジョン研究センター/研究会座長）

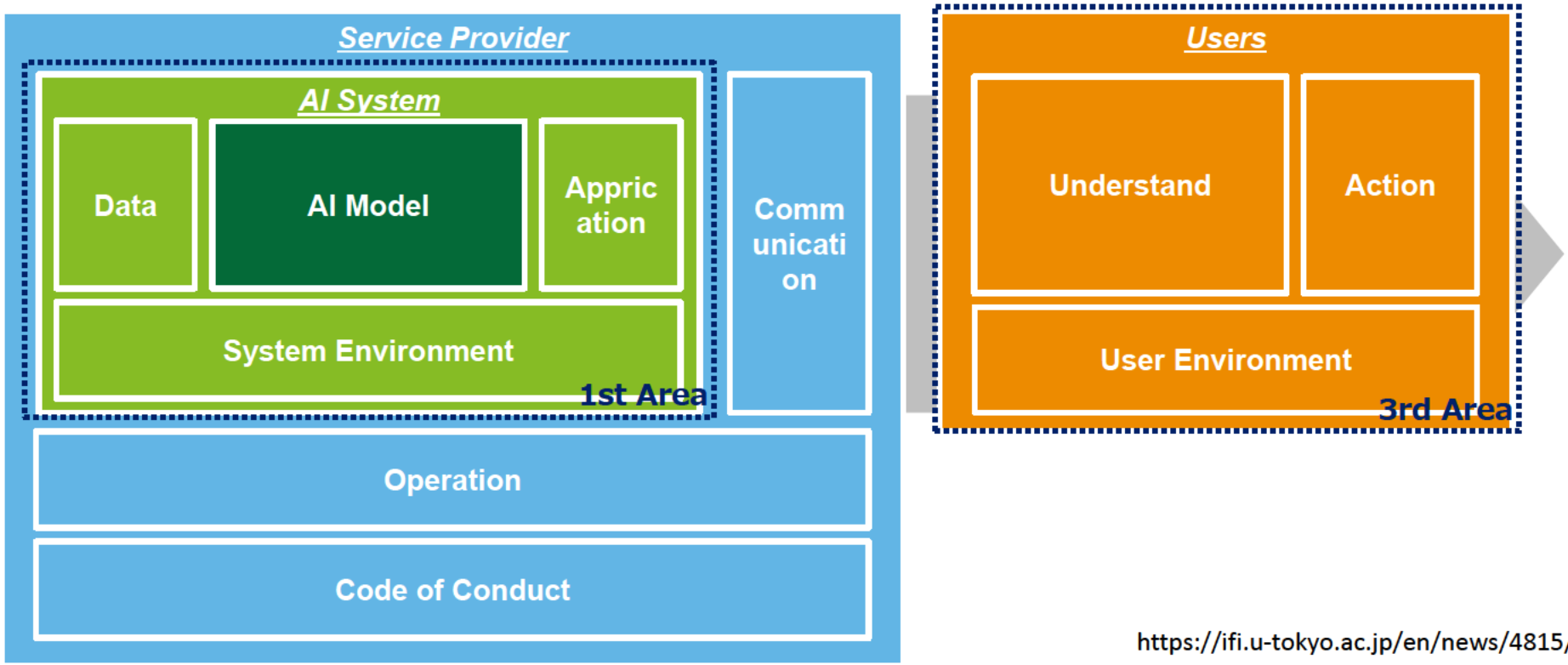
配布資料はこちら（JDLA会員限定）

開催レポート

https://www.jdla.org/about/studygroup/sg01/

Ex.2) Risk Chain Model

- Risk Management Tool for Companies / 企業のリスク管理ツール
- Collect Best Practices / ベストプラクティスの収集



https://ifi.u-tokyo.ac.jp/en/news/4815/

Companies
/ 企業

Corporate vision and values /
ビジョンや実現したい価値

↔ AI Principles

Consider AI Service Requirements &
Technologies / AIサービス要件と技術把握

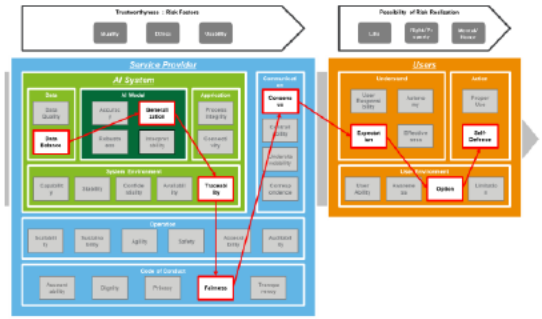
↔
- Facial Recognition
- Building Entrance

Create Risk Scenarios
/ リスクシナリオ作成

↔
- Data bias & wrong feedback
- Influence of the noise

Draw Risk Chains
/ リスク要因の関係性作成

↔



-Usage Policy
- Education
- Data modification

↔ Consider Risk Control
/ リスク管理実行

リスクアセスメント&コントロール (Case2：採用AI)
- リスクの複雑性評価 → リスクコントロールを検討 -

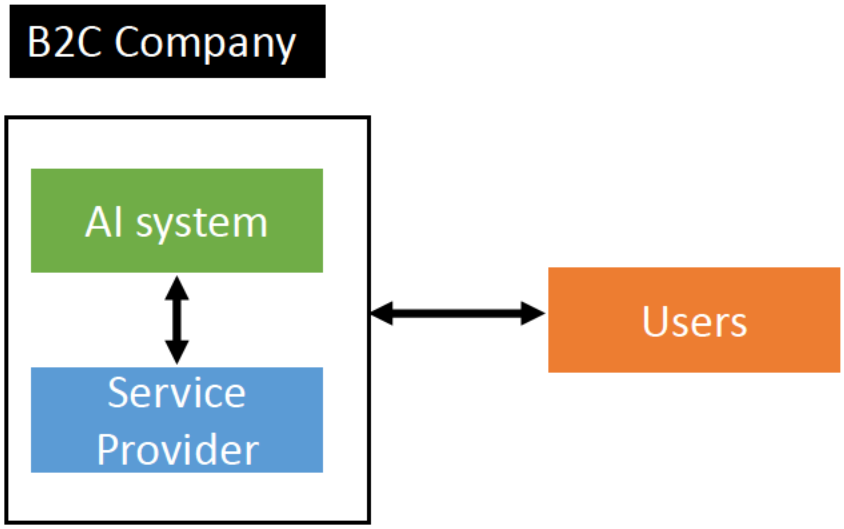
実現すべき価値・目的（リスクの影響先）		リスクNo.	リスクシナリオ	技術的難易度	環境変化	利用者起因	RC	コントロールのサマリー		
								AIシステム	サービスプロバイダ	ユーザー
1	人材採用レベルの維持・向上	R001	適切な評価	○			●	システム環境の確保 個別モデルの開発	目標精度の設定・見直し 再学習の指示	正確なフィードバック
		R002	予測性能の維持	○			●	十分な正解率の確保 検証可能性の確保	予測性能の検証 再学習	代替運用
		R003	ノイズによる影響	○			●	モデルの頑健性 判断根拠の出力	判断根拠の分かりやすさ 判断根拠の検証	判断根拠の検討
		R004	虚偽の申込	○			●	十分な正解率の確保 判断根拠の出力	過去の異常事例の検証 利用部門との連携	最終選考プロセス
		R005	過度なAI依存	○		○	●	判断根拠の出力	判断精度の開示 判断根拠の分かりやすさ	予測性能・リスクの認識 最終判断プロセス
		R006	誤ったフィードバック	○		○	●	学習データの検証 学習時の情報の保管	判断精度・異常値の検証 再学習	正確なフィードバック 人事システムからの反映
		R007	人材トレンドの変化	○	○		●	データ分布変化の認識 汎化性能の見直し	分布／正解率の監視 再学習	人材トレンド変化の認識
		R008	新たな職種	○	○		●	開発環境の準備 学習データ確保 モデルの開発	開発体制の整備 開発したモデルの検証	要求仕様の定義
2	採用活動に係るコストの削減	R009	コスト超過					適切な価格設定	コスト管理	
3	海外グループを含めたサービス提供	R010	地域の会社への対応	○	○		●	システム環境の確保 個別モデルの開発	個別の目標精度の設定 モデルの性能監視 開発体制の確保	
		R011	不十分な開発スピード					再学習環境の準備 モデル開発者の確保	PJ体制の確保	
4	企業の社会的責任(公平性のある採用活動)	R012	判断根拠情報の不正販売	○		○	●	操作ログの記録	不正利用の監視 外部弁護士との連携	内部牽制
		R013	公平性	○		○	●	データの偏り モデルの汎化性	公平性ポリシーの検討 ネガティブな判断の開示	AIの判断傾向の理解 公平な最終判断
		R014	予測結果の目的外利用			○		データ保護	アクセス管理 目的内利用	データの取扱
		R015	風評被害			○		データ保護	職業倫理の教育	職業倫理の教育
		R016	プライバシー保護			○		データ保護	法令順守の教育	法令順守の教育 データの取扱

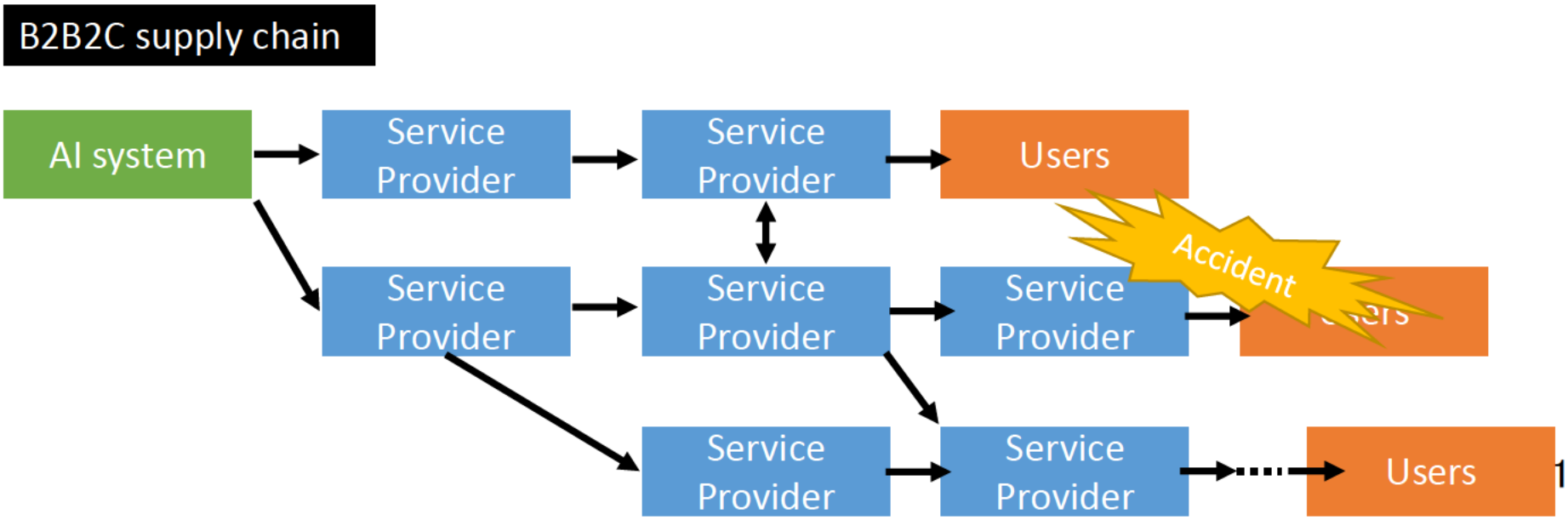
ケーススタディ

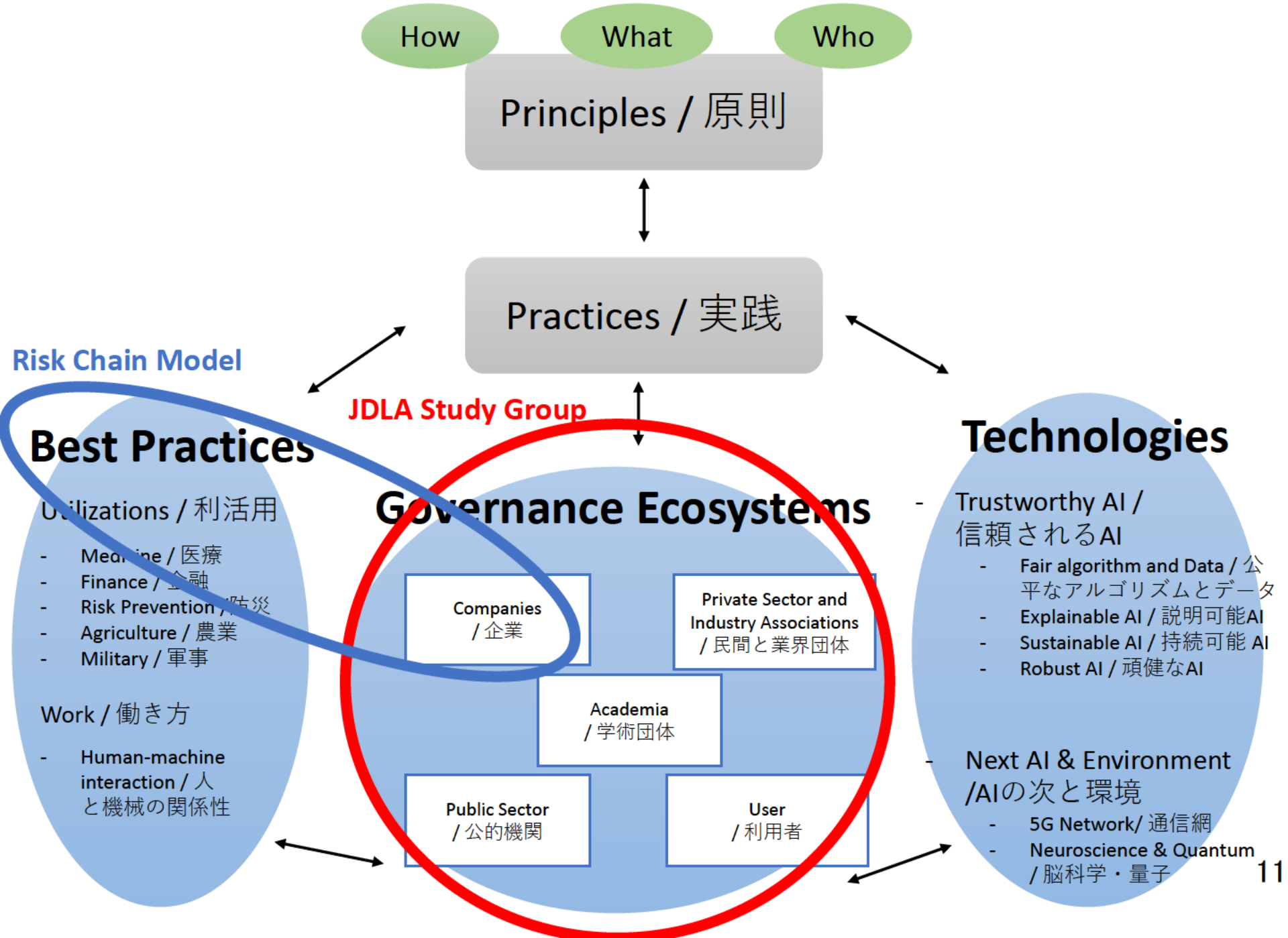
No	ケース	ジャンル
1	無人コンビニ	小売
2	採用AI	人事
3	送電線の点検AI	インフラ
4	不良品検知AI	製造
5	道案内ロボット	生活
6	がん診断AI	医療
7	ローン審査AI	金融
8	再犯可能性の検証AI	警備
9	無人コールセンター（チャットボット）	サービス
10	危険運転の検知AI	自動車
11	無人バス	交通
12	自動記事の作成	メディア
13	スポーツ採点AI	スポーツ
14	AIスピーカー	生活
15	自動運転（レベル３）	自動車

No	ケース	ジャンル
16	万引き防止AI	警備
17	トンネル点検AI	交通
18	裁判員AI	公共
19	食品の新商品開発AI	食品
20	無人ショベルカー	建設
21	エネルギー最適化AI	インフラ
22	教育カリキュラム作成AI	教育
23	農作物監視AI	農業
24	顔認証による自動決済	金融
25	熟練工の擬人化	製造
26	人事評価AI	人事
27	危険物品の検知	物流
28	AI自動運転（レベル５）	自動車
29	災害予測AI	インフラ
30	政策提言のAI	公共

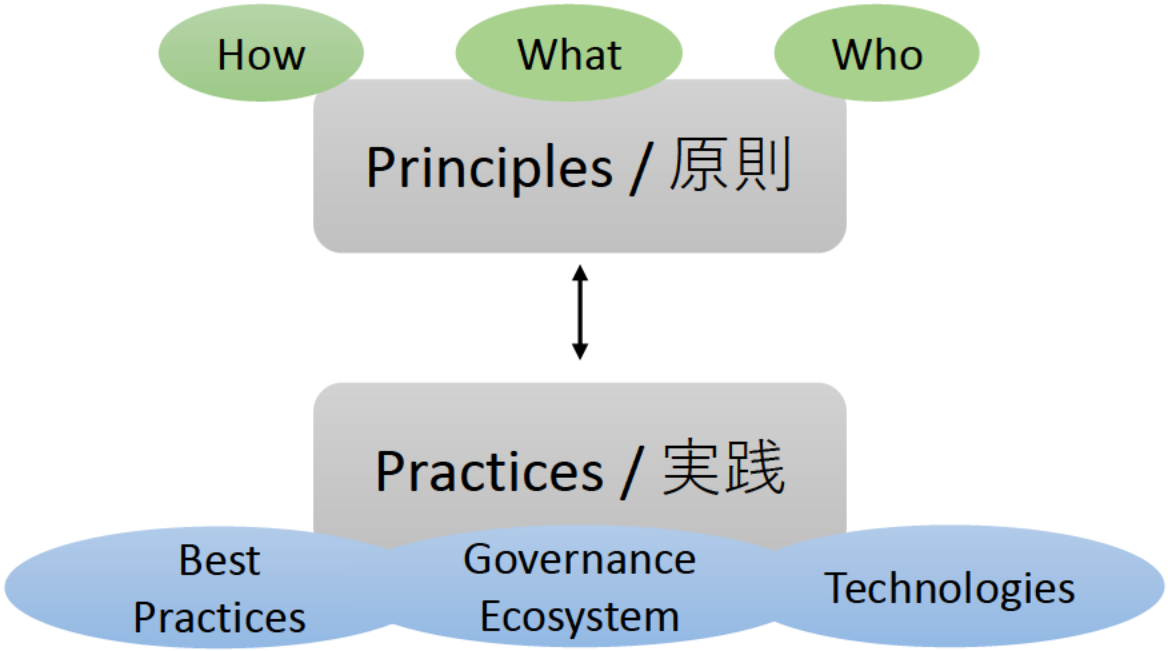
Industrial Structure






How
What
Who
Principles / 原則
Practices / 実践
Risk Chain Model
JDLA Study Group
Best Practices
Utilizations / 利活用
- Medicine / 医療
- Finance / 金融
- Risk Prevention / 防災
- Agriculture / 農業
- Military / 軍事
Work / 働き方
- Human-machine interaction / 人と機械の関係性
Governance Ecosystems
Companies / 企業
Private Sector and Industry Associations / 民間と業界団体
Academia / 学術団体
Public Sector / 公的機関
User / 利用者
Technologies
- Trustworthy AI / 信頼されるAI
  - Fair algorithm and Data / 公平なアルゴリズムとデータ
  - Explainable AI / 説明可能AI
  - Sustainable AI / 持続可能 AI
  - Robust AI / 頑健なAI
- Next AI & Environment /AIの次と環境
  - 5G Network/ 通信網
  - Neuroscience & Quantum / 脳科学・量子

- More Diverse and Inclusive stakeholders / より多様・包摂的な関係者に
- Aware this is cooperative and competitive areas
  / 競争かつ協調領域との意識を持つ



- Encourage and support stockholders to move on to practices
  / 各ステークホルダーに実践に移れるよう支援
- Store cases best practices and share within Japan and Abroad
  / 国内外への事例共有と発信