パネルディスカッション

パネルディスカッション
「AIセキュリティ」
その脅威と対策を考える

JNSA

議題（AIセキュリティについて）

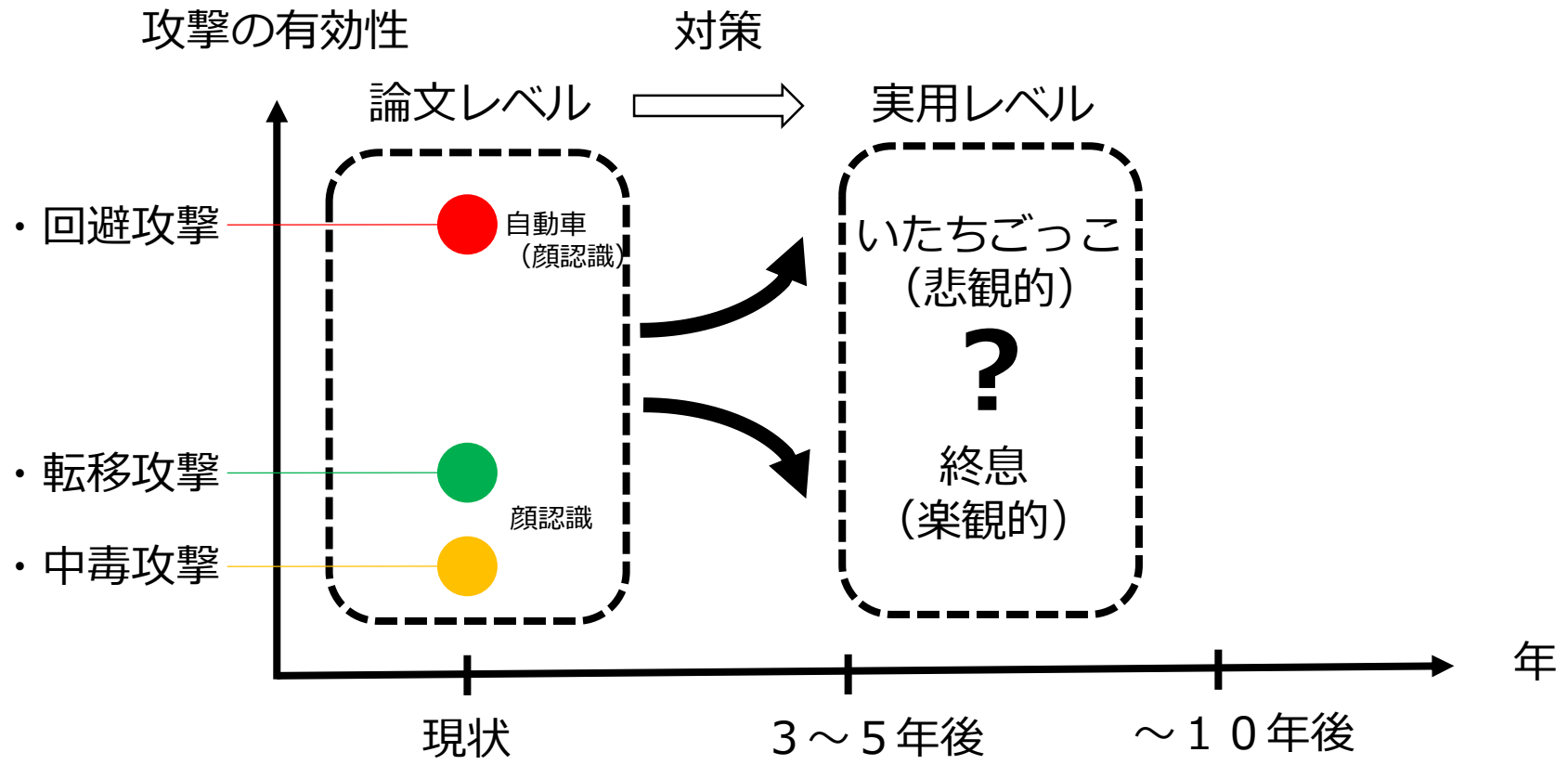
1．AIでのセキュリティ対策（～近い将来、～課題）

2．AIによる攻撃（～近い将来、～課題）

3．AI脅威論とセキュリティについて

（10年後、セキュリティ対策・運用の現場はどうなるか？）

1．AIでのセキュリティ対策（～近い将来、～課題）
セキュリティ対策の将来予測（仮説）



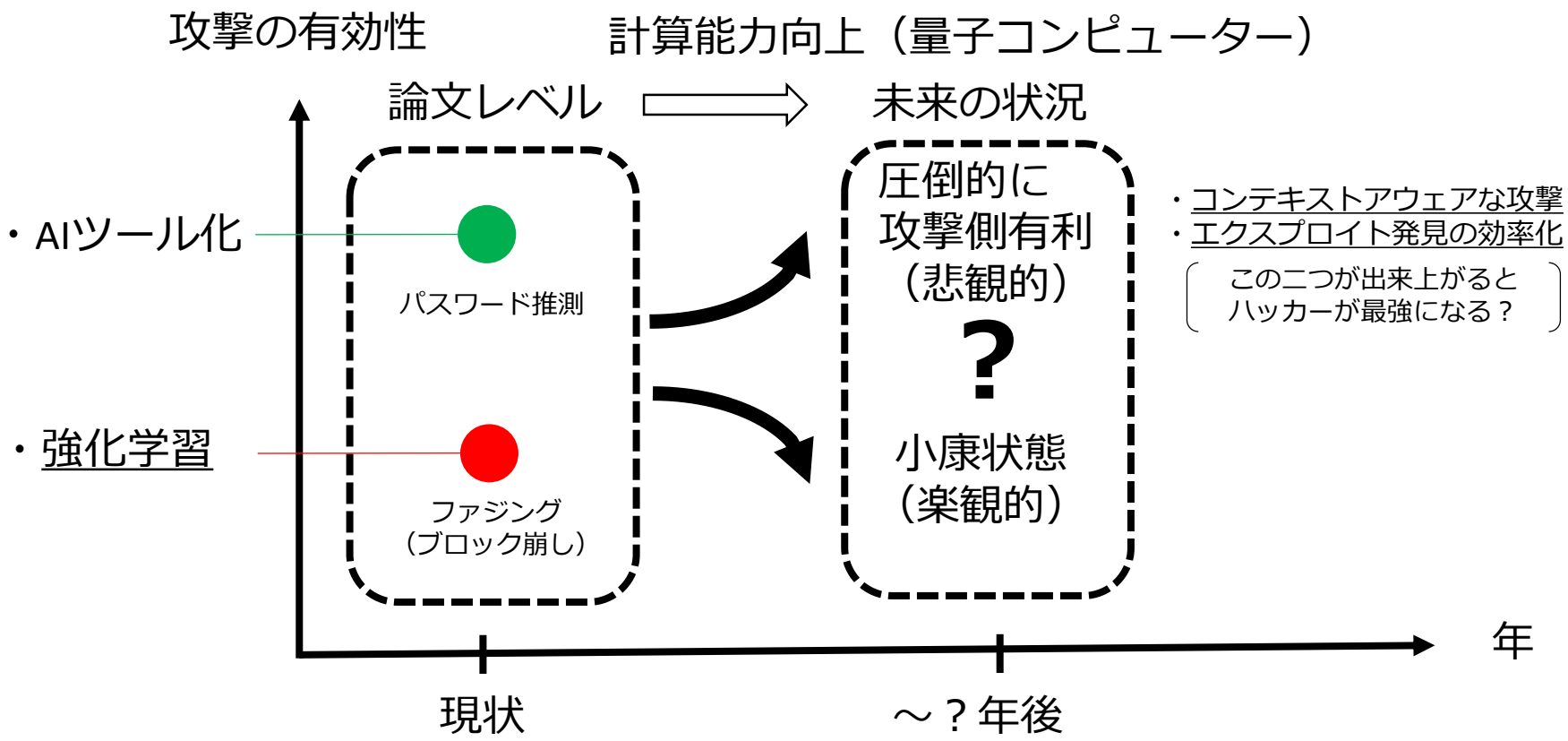
1．AIでのセキュリティ対策（～近い将来、～課題）
セキュリティ対策の将来予測（仮説）

脅威	現状の主となる対策	近い将来の対策（仮説）
回避攻撃	・敵対的訓練 ・ロジックでの検知	・ニューラルネットの堅牢化 （増大、再構成、逆伝搬等）？
中毒攻撃	・異常値検出	異常検知で防げる？
移転攻撃	・匿名化 ・アクセス制限	プライバシー保護で防げる？

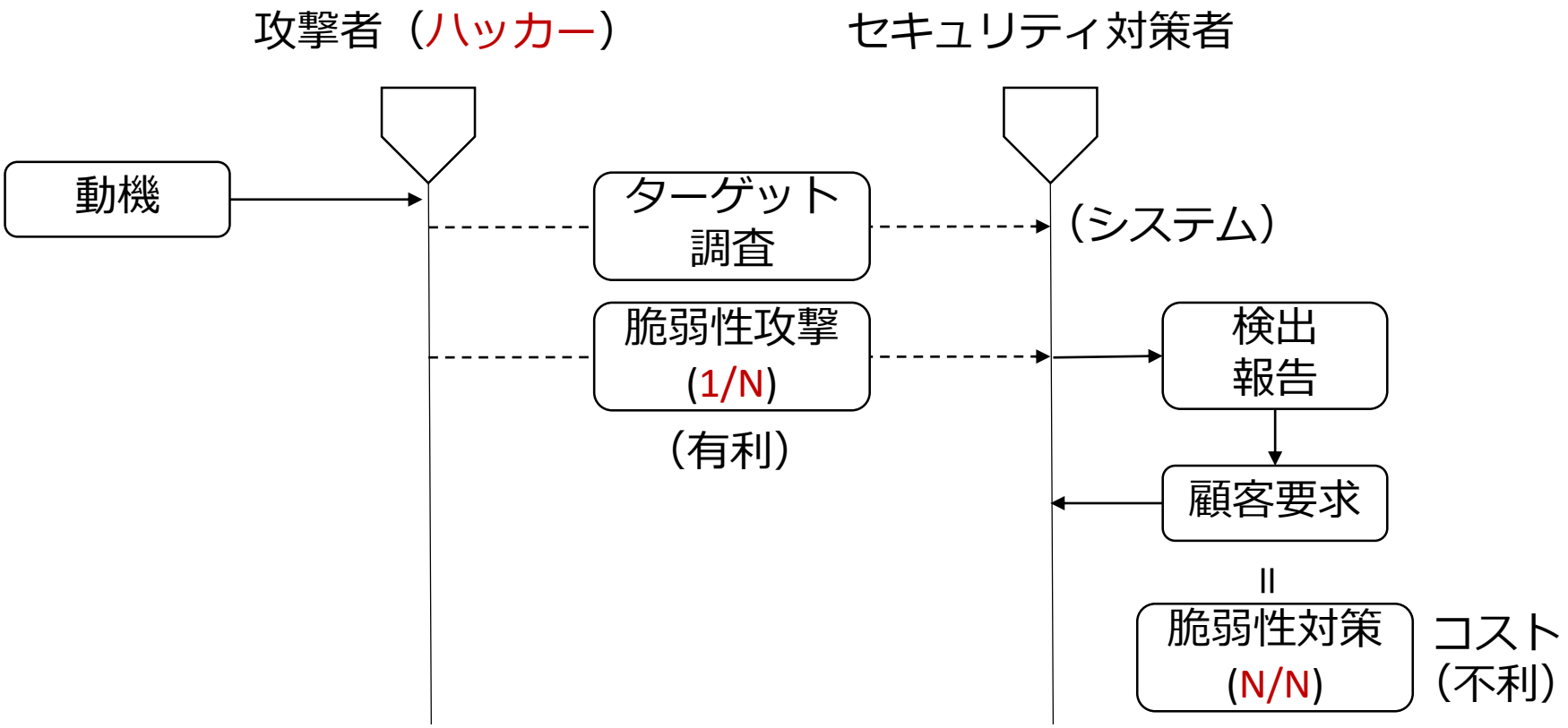
（将来見通し）
- 回避攻撃はニューラルネットの増強・改造で乗り切れる？
- 中毒攻撃は検知手段と攻撃の「いたちごっこ」？
  （再学習段階の「ビヘイビア」監視でも検知できる？）
- 移転攻撃は「プライバシー保護」などの対策で抑え込める？

①機械学習自体での改善？　②セキュリティ技術に頼る？

2．AIによる攻撃（可能性）

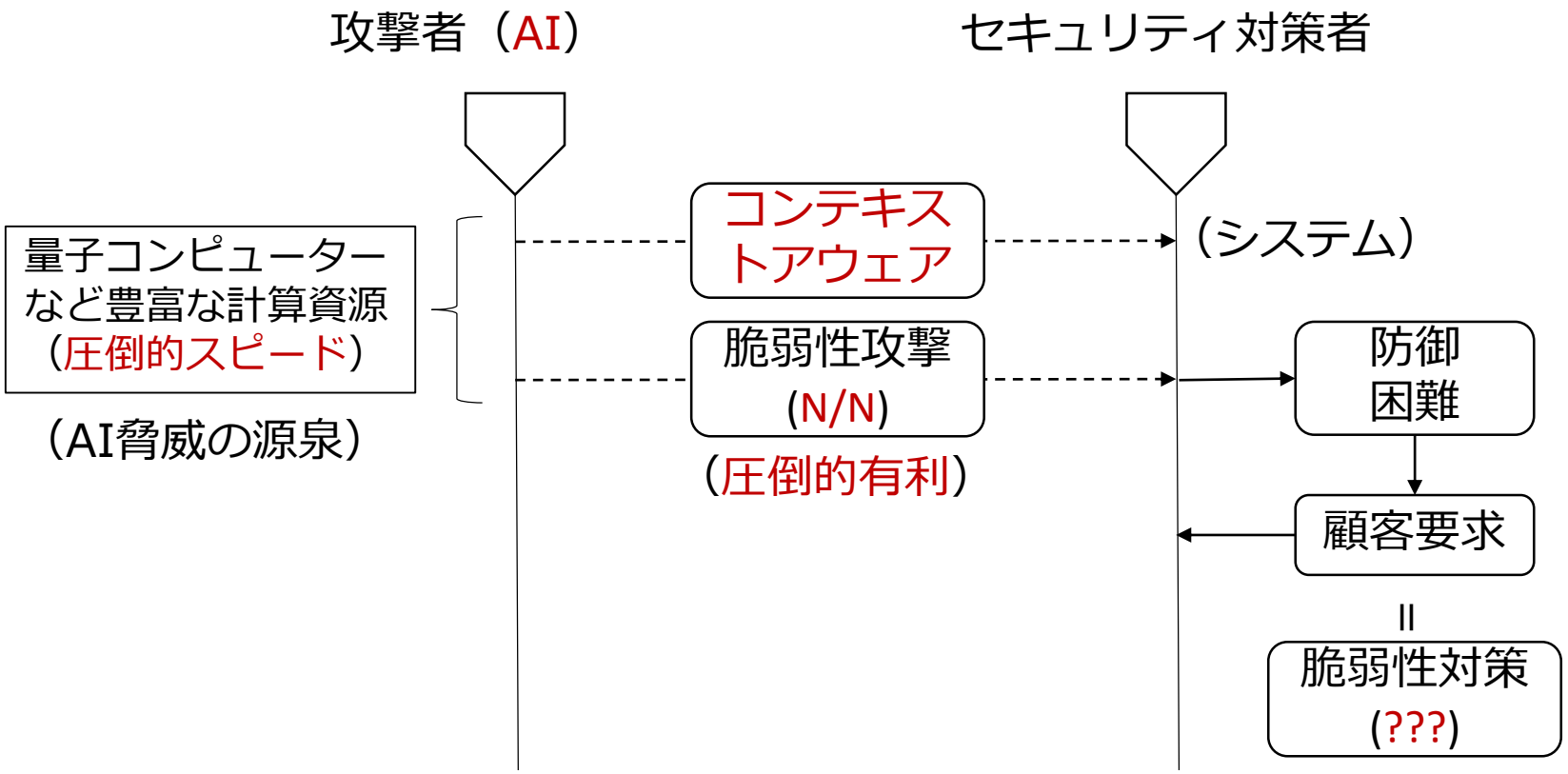


3. AI脅威論とセキュリティについて（現状）
（10年後、セキュリティ対策・運用の現場はどうなるか？）



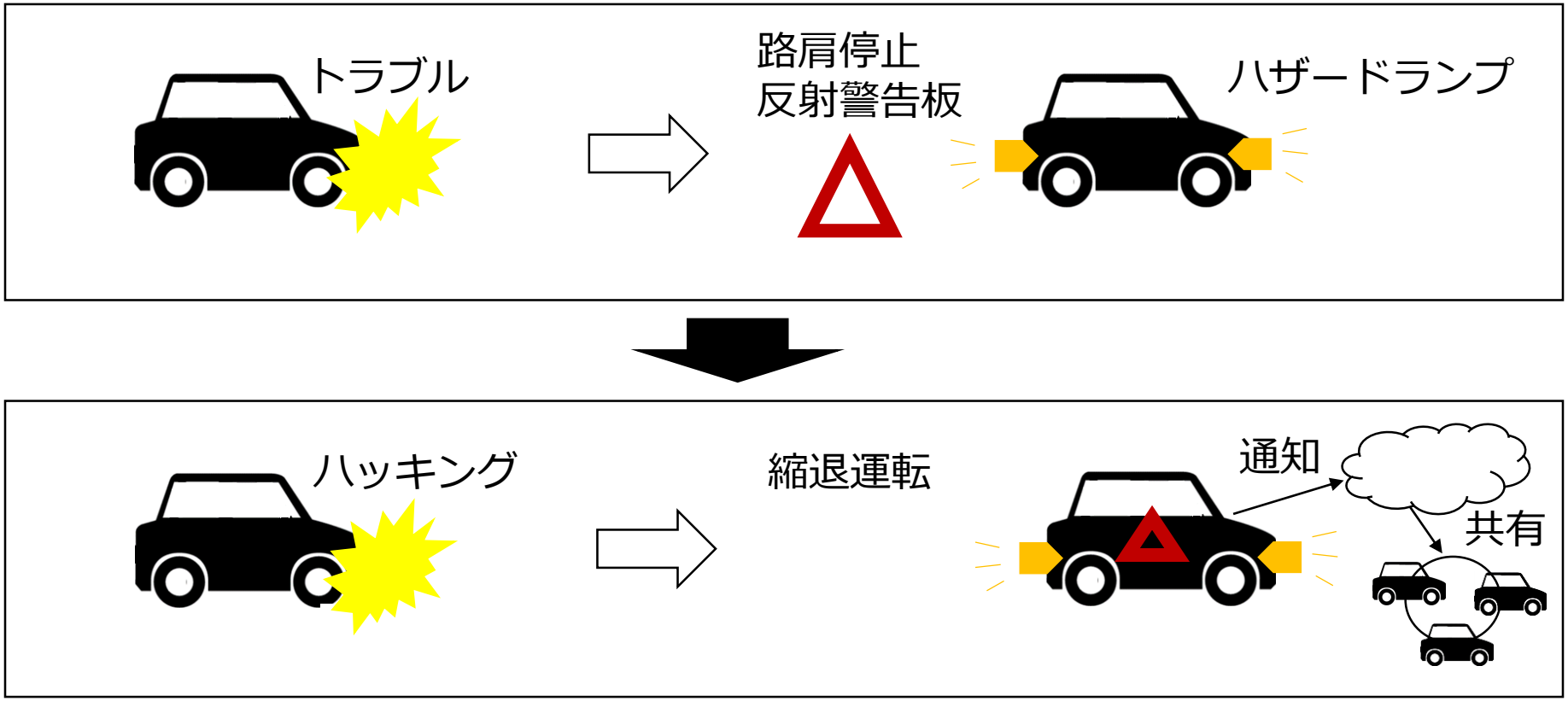
網羅性の要求→AIではなく説明できる「ルール」「ロジック」を採用する傾向

3. AI脅威論とセキュリティについて（将来）
（10年後、セキュリティ対策・運用の現場はどうなるか？）



「セキュリティホールを潰す」という手法は限界？（課題？）

3. AI脅威論とセキュリティについて（対策？）
（10年後、セキュリティ対策・運用の現場はどうなるか？）



①攻撃に対する堅牢性（縮退運転） ②攻撃通知（情報共有）