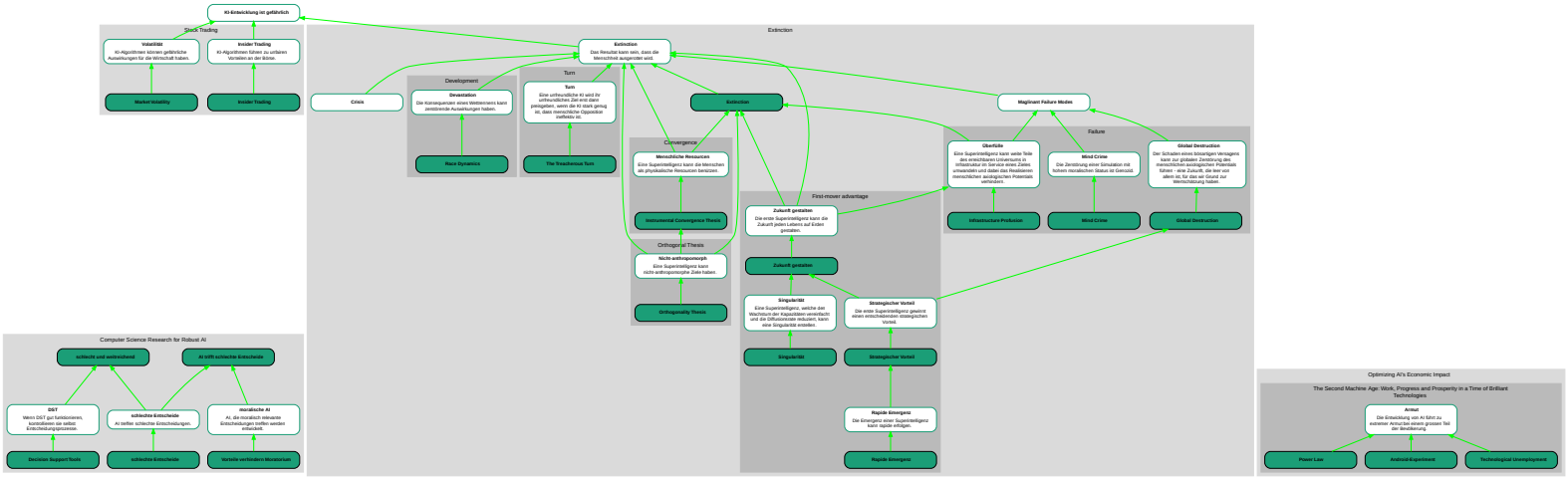
KI-Entwicklung ist gefährlich
Stock Trading
Volatilität
KI-Algorithmen können gefährliche Auswirkungen für die Wirtschaft haben.
Insider Trading
KI-Algorithmen führen zu unfairen Vorteilen an der Börse.
Market Volatility
Insider Trading
Extinction
Extinction
Das Resultat kann sein, dass die Menschheit ausgerottet wird.
Crisis
Development
Devastation
Die Konsequenzen eines Wettrennens kann zerstörende Auswirkungen haben.
Race Dynamics
Turn
Turn
Eine unfreundliche KI wird ihr unfreundliches Ziel erst dann preisgeben, wenn die KI stark genug ist, dass menschliche Opposition ineffektiv ist.
The Treacherous Turn
Extinction
Convergence
Menschliche Resourcen
Eine Superintelligenz kann die Menschen als physikalische Ressourcen benützen.
Instrumental Convergence Thesis
Orthogonal Thesis
Nicht-anthropomorph
Eine Superintelligenz kann nicht-anthropomorphe Ziele haben.
Orthogonality Thesis
First-mover advantage
Zukunft gestalten
Die erste Superintelligenz kann die Zukunft jeden Lebens auf Erden gestalten.
Zukunft gestalten
Singularität
Eine Superintelligenz, welche der Menschheit überlegen ist, wird sich rasch und die Öffentlichkeit reduziert, kann eine Singularität entstehen.
Strategischer Vorteil
Die erste Superintelligenz gewinnt einen entscheidenden strategischen Vorteil.
Singularität
Strategischer Vorteil
Rapide Emergenz
Die Emergenz einer Superintelligenz kann rapide erfolgen.
Rapide Emergenz
Maginant Failure Modes
Failure
Überfülle
Eine Superintelligenz kann verlor fette Ziele, welche Ressourcen und Infrastruktur in Sie sich eines Zieles, umwandeln und dabei das Realisieren menschlichen ausgiebigen Potentials verhindern.
Infrastructure Profusion
Mind Crime
Mind Crime
Die Zerstörung einer Simulation mit hohem moralischen Status ist Genozid.
Global Destruction
Der Schaden eines bösartigen Versagens kann die Zerstörung des gesamten menschlichen ausgiebigen Potentials führen – eine Zukunft, die leer von allem ist, für das wir Grund zur Wertschätzung haben.
Global Destruction
Computer Science Research for Robust AI
schlecht und weitreichend
AI trifft schlechte Entscheide
DST
Wenn DST gut funktioniert, kann man sie selbst Entscheidungsprozesse.
schlechte Entscheide
AI trifft weniger schlechte Entscheidungen.
moralische AI
AI, die moralisch relevante Entscheidungen treffen werden entwickelt.
Decision Support Tools
schlechte Entscheide
durch verfeinertes Monitoring
Optimizing AI's Economic Impact
The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies
Armut
Die Einführung von KI führt zu extremer Verarmung grosser Teil der Bevölkerung
Power Law
Android-Expansion
Technological Unemployment