Superintelligence as existential risk

Control Problem Argument
Motivation selection is the only control method that has a chance to avoid the doomsday take-over scenario (cf. Chapter 9).

Value-based control methods required

Superintelligence as serious possibility
Given alternative technological realizations and the positive feedbacks involved in the development of intelligent systems, the creation of a superintelligent system within this century is a serious possibility (cf. Chapters 2-5).

Bootstrapping
Intelligent systems can (be used to) further improve their cognitive abilities, creating positive feedbacks in SI development (cf. Chapters 4,5).

Superpower
Take-over scenario argument
A superintelligence system is likely to bring under control all resources on Earth, reducing humans to mere means to obtain non-anthropomorphic final goals, and thus leading to the extinction of human civilization as we know it (cf. Chapter 8).

Orthogonality thesis

instrumental convergence thesis

No societal preparation without R&D
Only once R&D efforts into SI are underway, will society start to prepare seriously for the advent of superintelligence; so the earlier R&D starts, the more time we have to get ready (cf. pp. 293-4).

Control methods
Intelligent systems can be controlled either through capability containment or motivation selection (cf. Chapter 9).

Capability control ineffective
Capability control is ineffective and unreliable for superintelligent systems, because the SI is likely to see ways to manipulate the world which we don't (cf. Chapters 9, 12, 13).

Superintelligence as serious possibility

Slower SI development!

The principle of differential technological development
The development of cognitive enhancement poses existential risks, namely inasmuch as it represents a path to superintelligence.

Cognitive enhancement development too risky
The development of cognitive enhancement reduces an existential risk, namely inasmuch as it prepares us for a superintelligence explosion – specifically by enabling us tackle the control problem (cf. pp. 288-90).

Cognitive enhancement prepared us for SI
The development of cognitive enhancement reduces an existential risk, namely inasmuch as it prepares us for a superintelligence explosion (cf. pp. 288-90).

Slower cognitive enhancement development!

Machine SI development too risky
The development of machine SI poses existential risks.

Alternative paths to SI
There exist alternative technological paths to superintelligent systems (cf. Chapter 2).