

OBSAH

Předmluva 7

- KAPITOLA 1. **Pokud uspějeme** 9
Jak jsme sem došli? 11
Co se stane dál? 13
Co se pokazilo? 15
Dokážeme to opravit? 16
- KAPITOLA 2. **Intelligence u lidí a strojí** 19
Intelligence 19
Počítače 32
Inteligentní počítače 38
- KAPITOLA 3. **Jaký vývoj AI čeká v budoucnu?** 55
Blízká budoucnost 55
Kdy přijde superinteligentní AI? 65
Konceptuální průlomky před námi 66
Představy o superinteligentním stroji 77
Meze superintelligence 79
Jak bude AI prospěšná pro člověka? 81
- KAPITOLA 4. **Zneužití AI** 85
Stálý dohled, přesvědčování a ovládání 85
Smrtící autonomní zbraně 90
Likvidace práce v podobě, jak ji známe 92
Uchvácení dalších lidských rolí 100
- KAPITOLA 5. **Příliš inteligentní AI** 107
Problém gorily 107
Problém krále Midase 110
Strach a chamtivost: instrumentální cíle 113
Intelligenční exploze 114
- KAPITOLA 6. **Nikterak velkolepá debata o AI** 117
Popření 118
Odchylování 123
Tribalismus 126
Nemůžeme prostě... 127
Znovunastartovaná debata 134

KAPITOLA 7.	AI: odlišný přístup	135
	Principy pro prospěšné stroje	136
	Důvody k optimismu	140
	Důvod k opatrnosti	142
KAPITOLA 8.	Dokazatelně přínosná AI	145
	Matematické záruky	145
	Učit se preference z chování	149
	Asistenční hry	151
	Žádosti a pokyny	159
	Zadrátování hlavy	161
	Rekurzivní sebezdokonalování	163
KAPITOLA 9.	Komplikace: my	165
	Různí lidé	165
	Mnoho lidí	166
	Hodní, protivní a závistiví lidé	176
	Hloupí, emocionální lidé	179
	Mají lidé opravdu preference?	181
KAPITOLA 10.	Vyřešený problém?	189
	Prospěšné stroje	189
	Ovládání AI	191
	Zneužití	193
	Ochablost a lidská autonomie	194
PŘÍLOHA A.	Hledání řešení	197
	Rezignace na racionální rozhodování	198
	Pohled do budoucna	201
PŘÍLOHA B.	Znalosti a logika	205
	Výroková logika	205
	Logika prvního řádu	207
PŘÍLOHA C.	Nejistota a pravděpodobnost	209
PŘÍLOHA D.	Učení se ze zkušeností	218
	<i>Poděkování</i>	226
	<i>Poznámky</i>	227
	<i>Zdroje ilustrací</i>	259
	<i>Rejstřík</i>	261