

## Doktorandi individuaalplaani õppeaastateks 2007/2008 - 2011/2012

**Doktorant:** Roland Pihlakas

**Õppekava, teaduskond/instituut; doktorikool:** Tehnika ja tehnoloogia, Tehnoloogiainstituut. Doktorikool on IKT DK, alates hetkest, mil sinna vastu võetakse.

**Doktoritöö teema:** Tehisintellekti turvalisus läbi mittepööratavate tegevuste vältimise.

**Juhendaja:** Maarja Kruusmaa

### **I. Õpingud (õppeained, AP-d, k.a. õppetöö läbiviimine doktoriõpingute raames; eraldi teistest ülikoolidest võetud õppeained).**

Õppeained: peamiselt pideva ja diskreetse matemaatika kursused. Täpsem loetelu lisas.

### **II. Uurimistöö (sh teadustöö, konverentsiettekanded, seminarid, publikatsioonid jm; hinnangulised töö tulemused ja maht AP-des).**

*Individuaalplaani lisas esitada põhjalik uurimistöö kava koos ajatabeliga, probleemiasetusega, põhjenduste, hüpoteeside ja kirjanduse loeteluga.*

Esimene aasta on peamiselt uurimistöö teema tausta uurimiseks ja teadmiste täiendamiseks. Sellega seoses plaanis võtta punktis üks mainitud kursuseid, lugeda artikleid ning raamatuid. Täpsemalt vaata lisast. Kui Tehnoloogiainstituudis hakkab see toimuma, siis lisaks ka teadusseminar.

### **III. Muud doktoritööga seotud ülesanded ja sellega seotud infrastruktuuri kasutamine (sh grantide ja teadusteemade täitmine, koostööprojektid jm).**

Ei ole veel.

### **IV. Muu (õppetöö läbiviimine, juhendamine, ülikoolisisestes ja -välistes erialaga seotud komisjonides ja projektides osalemine jm).**

Ei ole. Kui tekiks võimalus õppetöö läbiviimiseks, siis tõenäoliselt kasutaksin seda.

### **Doktorant**

..... 20. sept 2007

*allkiri, kuupäev*

### **Kinnitan:**

### **Juhendaja**

..... 20. sept 2007

*allkiri, kuupäev*

## **Lisa. Järgmise õppeaasta täpsustatud õpingute ja uurimistöö kava.**

Esimese õppeaasta eesmärk on süvendada taustateadmisi ning viia läbi esimesed katsed uuritava printsiibi rakendatavuse osas.

- **Raamatud ja artiklid**
  - 100 artiklit Developmental Robotics & Epigenetic Robotics konverentsidelt.
  - Leida ning lugeda varasemaid uurimistöö põhiteemaga, tehisintellekti turvalisusega, seonduvaid artikleid.
  - Raamat "Safe and Sound artificial intelligence".
  - Närvivõrkude raamat "Neural Networks, A Comprehensive Foundation" / S. Haykin.
  - BDI-agendid (belief-desire-intention agendid).
  - Raamat "Mis asi see on, mida nimetatakse teaduseks?" / A. F. Chalmers.
  - Raamat "Research methods for postgraduates" / T. Greenfield.
  - Mõtlemise mudeliga seoses uurida Perceptual Control Theory mudelit, mille võib olla mõningat ühisosa.
  - Muud kontrolliteooriad.
  - Automaattõestajad ning loogiline programmeerimine.
- **Katsed (nendega tegelemise ulatus sõltub kevadsemestri tunniplaani, mida pole veel võimalik kindlaks määrata):**
  - Senise naturaalse mõtlemise mudeli katsetused mittepööratavuse rakenduses.
  - Arendada naturaalse mõtlemise mudelit edasi või täpsemini – lisada sobiv karkass mõtleva-õppiva osa ümber, et esindada mittepööratavusi ja eesmärgi vastavalt uurimiseks oleva tehisintellekti ohutuse printsiipidele.
  - Võrrelda mittepööratavuse üht spetsiifilist rakendust – "kokkupõrgete vältimist". Võrrelda selle naturaalse mõtlemise mudelil põhineva teostuse jõudlust teiste mudelitega, mis on loodud puhtalt kokkupõrgete vältimise eesmärgil.
  - Õiguste ja eesmärkide esitamine loogilise programmeerimise keeles.
- **Võimalikud artiklite teemad (nendega tegelemise ulatus sõltub kevadsemestri tunniplaani, mida pole veel võimalik kindlaks määrata):**
  - Blokeerimine klassikalises tingimises kui credit assignment üks lahendusi, võrdlused teiste õppimisalgoritmidega.
  - Eesmärgisüsteem, kus lisaks piirangutele-õigustele (mittepööratavuse vältimisele) on ka eesmärgid töösse lülitatud. See oleks edasiarenduseks senistest artiklitest uuritava printsiibi teemal.
  - Mitme piirangu ja mitme eesmärgi mõju süsteemile.
- **Kursused:**
  - Pidev ja diskreetne matemaatika, loogilise ja funktsionaalprogrammeerimise kursused.
  - Sügis 2007:
    - MTPM.06.030 Matemaatiline analüüs I (4 AP)
    - MTAT.05.072 Sissejuhatus matemaatilisse loogikasse (3 AP)
    - MTAT.05.047 Funktsionaalprogrammeerimise meetod (2 AP)
    - MTAT.06.008 Tehisintellekt I (4 AP)
    - MTRM.01.020 Hulgateooria ja matemaatiline loogika (2 AP)