Filozófiai útvonalak a mesterséges intelligencia emancipációjához

Héder Mihály (BME)

National Research, Development and Innovation Office
Hungary

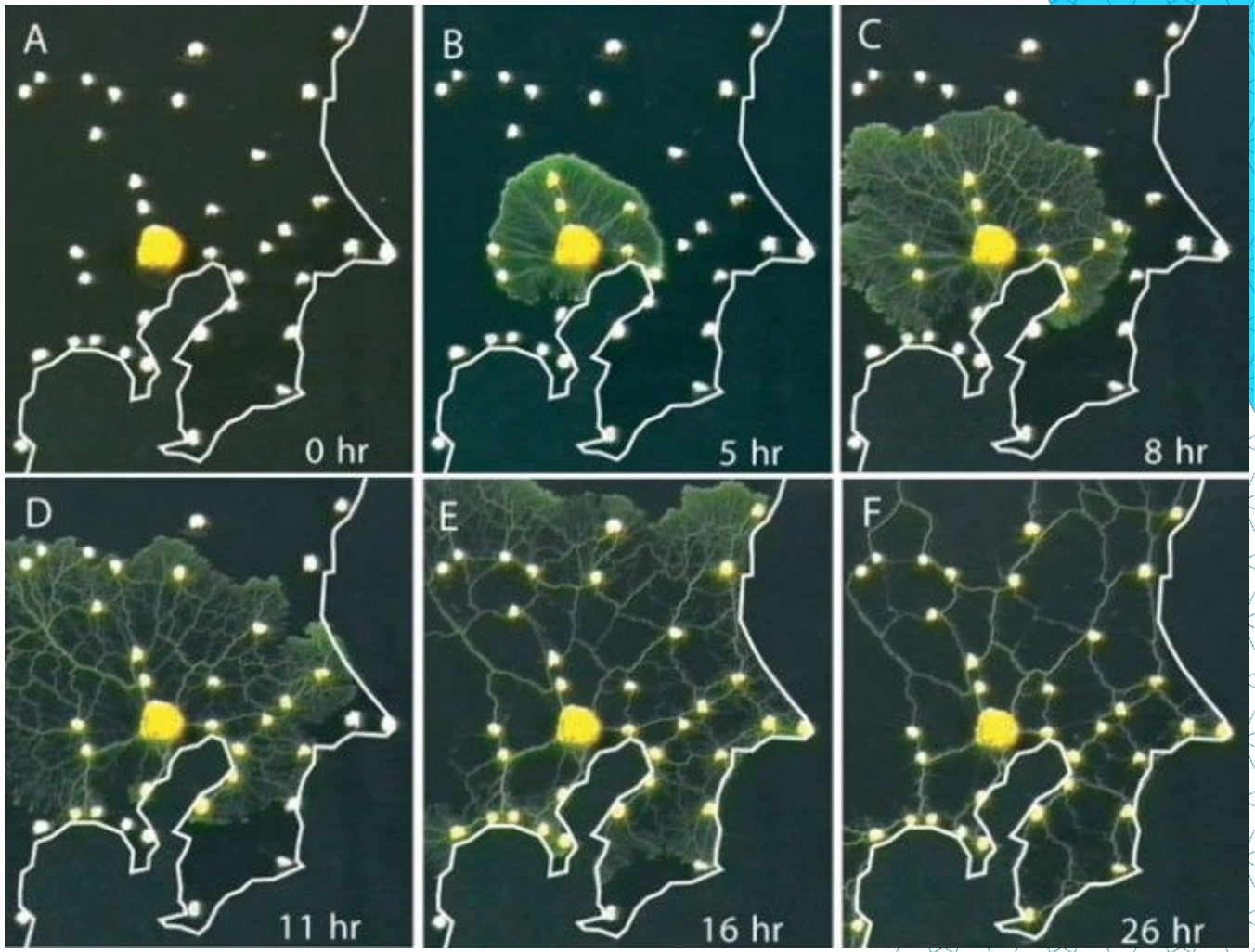
PROGRAM
FINANCED FROM
THE NRDI FUND

Ontológiai kérdések

Valóság? És Valóság!

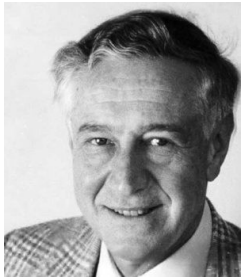
- Hol van a **Turing Gép**?
- Opciók
  - Hyperuranion
  - fizikalizmus
  - anti-realizmusok (az emberek fejében)
  - ??
- Tulajdonságok
  - tökéletesség
  - állandó hibamentesség
  - szélsőséges, sarkított kiterjedések
    - tökéletes üresség, végtelenség

Tokiói metró

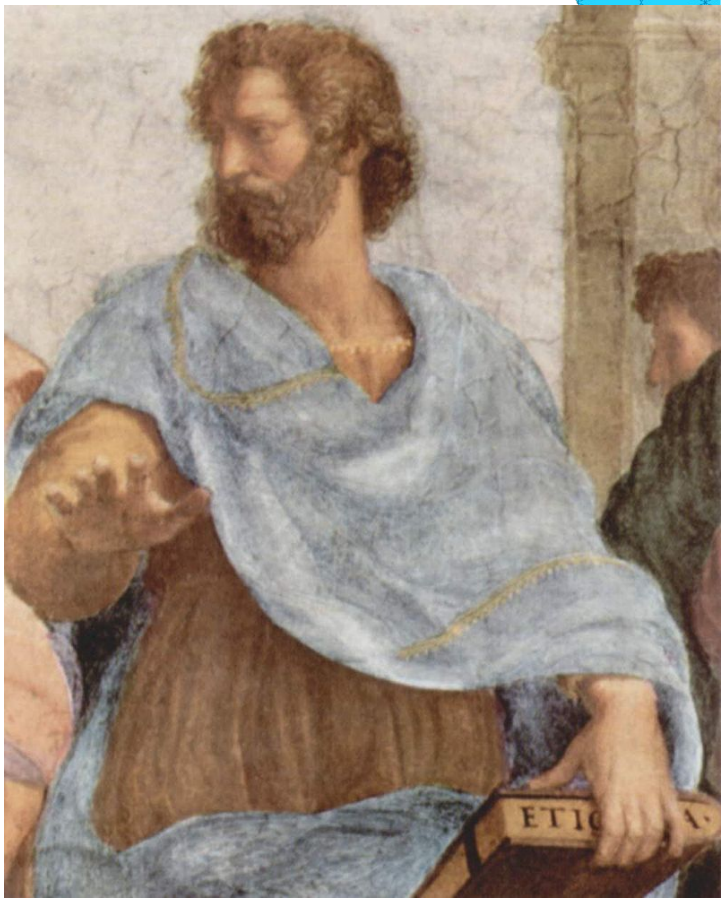


Valóság? És Valóság!

- Hol van a számítógép?
- Egyetlen opció
  - a természetben
- Tulajdonságok
  - végesség
  - tökéletlenség
  - átmenetiség
  - entrópia

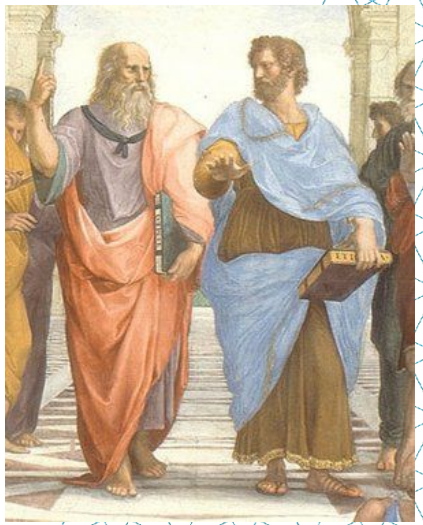


Rolf
Landauer
1927 - 1999



Inhibitor

- **Tézis:** *A megértés gátjainak forrása az, hogy a számítógép modelljének tulajdonságait keverjük a számítógép tulajdonságaival*
  - ***Még részletesebben:*** *a modell és a számítógép ontológiai státuszát össze nem egyeztetett metafizikai keretekből vezetjük le*
    - *nem lesz jó a morális státusz kérdéshez*
  - *Aluldeterminált, de nem úgy és nem annyira*



Rafaeló 1511,
Az athéni
akadémia

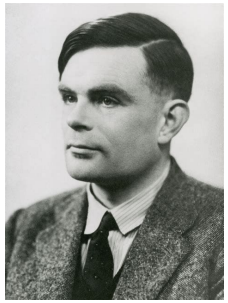
Következmények

Példa a kavargó világképre: CPS

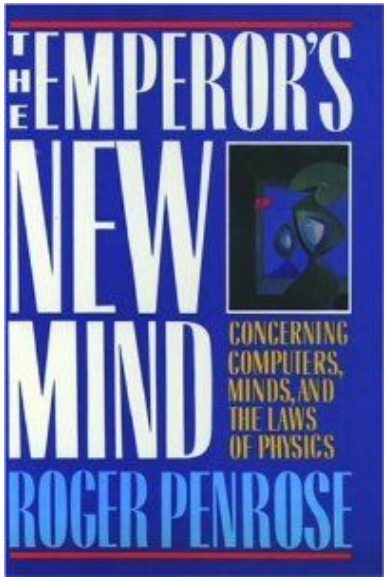
Egymással szoros összefonódásban működő hardver és szoftver komponensek, melyek különböző térbeli és időbeli skálákon operálnak, különböző viselkedési modelleket jelenítenek meg, és egymással rengeteg különböző módon képesek együttműködni, a változó kontextushoz adaptálódva. Kíber-fizikai rendszerként említhetjük az intelligens elektromos hálózatokat, az önvezető járműveket, az ipari folyamatok monitorozását, vagy az idősek otthoni felügyeletéért felelős megoldásokat. (forrás: HTE)

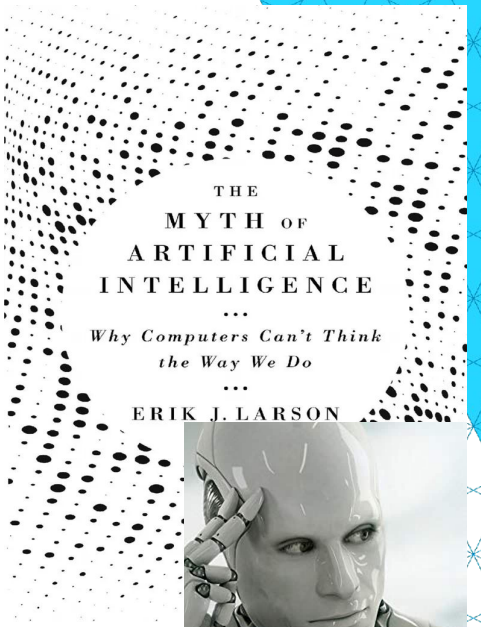
Gödel nemteljességi tételei

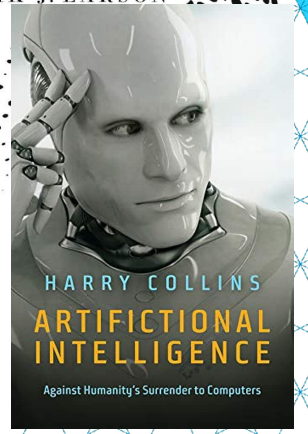
(3) *The Mathematical Objection.* There are a number of results of mathematical logic which can be used to show that there are limitations to the powers of discrete-state machines. The best known of these results is known as Gödel's theorem,[1] and shows that in any sufficiently powerful logical system statements can be formulated which can neither be proved nor disproved within the system, unless possibly the system itself is inconsistent. There are other, in some respects similar, results due to *Church*, *Kleene*, *Rosser*, and *Turing*. The latter result is the most convenient to consider, since it refers directly to machines, whereas the others can only be used in a comparatively indirect argument: for instance if Gödel's theorem is to be used we need in addition to have some means of describing logical systems in terms of machines, and machines in terms of logical systems. The result in



Chrisley, R. (2005) Simulation and Computability: Why Penrose fails to prove the impossibility of Artificial Intelligence (and why we should care), URL: http://www.idt.mdh.se/ECAP-2005/articles/COGNITION/RonChrisley/RonChrisley.pdf







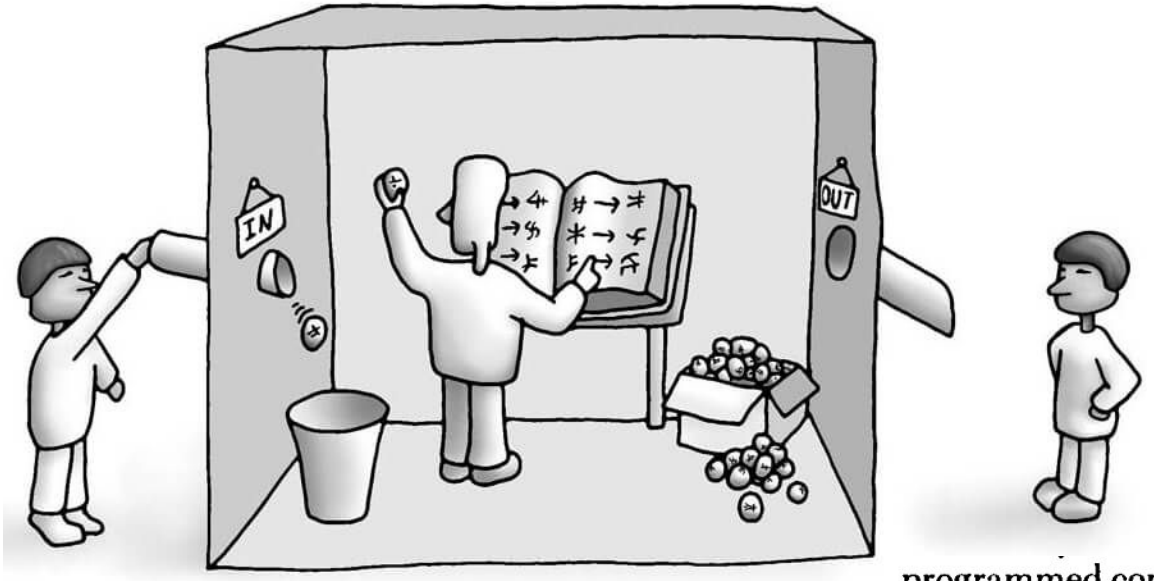
Gödel első nemteljességi tétele

Minden ellentmondásmentes, a természetes számok elméletét tartalmazó, formális-axiomatikus elméletben megfogalmazható olyan állítás, mely se nem bizonyítható, se nem cáfolható.

$$\mathrm{Consis}(\mathcal{T}) : \neg\Box\mathbf{f}$$

$$\mathcal{T} \vdash \mathrm{Consis}(\mathcal{T}) \rightarrow \mathbf{G}$$
$$\mathcal{T} \vdash \mathrm{Consis}(\mathcal{T}) \implies \mathcal{T} \vdash \mathbf{f}$$

Rossz érvek 2. szintaxis vs. szemantika vs LLM-ek



programmed computer does not do "information processing." Rather, what it does is manipulate formal symbols. The fact that the programmer and the interpreter of the computer output use the symbols to stand for objects in the world is totally beyond the scope of the computer. The computer, to repeat, has a syntax but no semantics. Thus, if you type into

Is there any syntax in the computer?

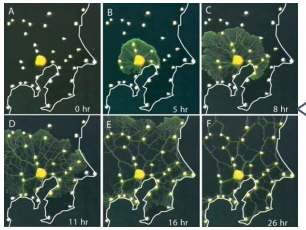
Because the formal symbol manipulations by themselves don't have any intentionality; they are quite meaningless; they aren't even *symbol* manipulations, since the symbols don't symbolize anything. In the linguistic jargon, they have only a syntax but no semantics. Such intentionality as computers appear to have is solely in the minds of those who program them and those who use them, those who send in the input and those who interpret the output.

The aim of the Chinese room example was to try to show this by showing that as soon as we put something into the system that really does have intentionality (a man), and we program him with the formal program, you can see that the formal program carries no additional intentionality. It adds nothing, for example, to a man's ability to understand Chinese.

Mire jók a modellek?

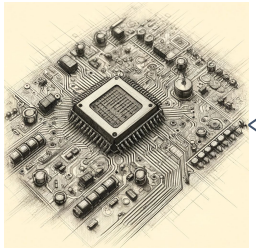
Model vs Valóság

- Jó modell: prediktív
- Szinte sosem teljes
- Szinte sosem teljesen pontos



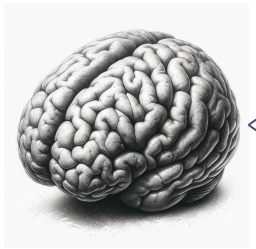
Optimalizáló számítás

Kalorikus gép



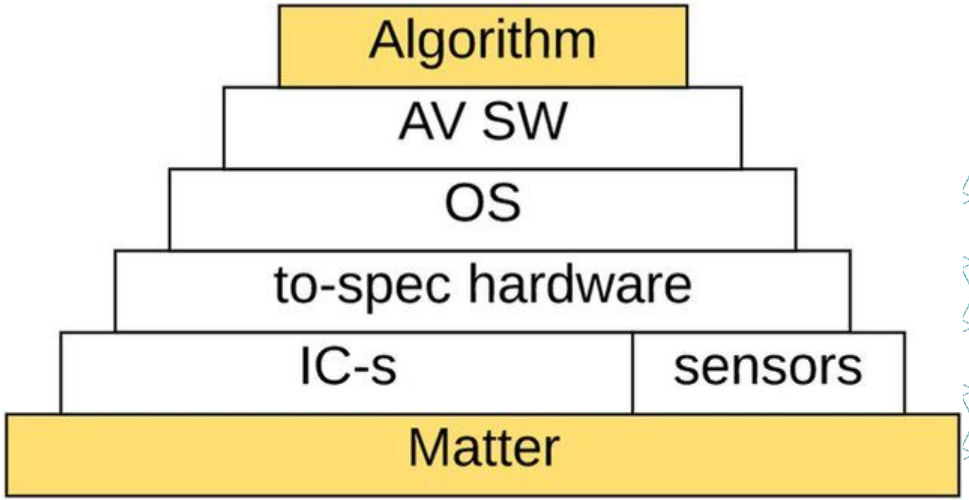
Véges Automata

Kalorikus gép



Véges Automata

Kalorikus gép

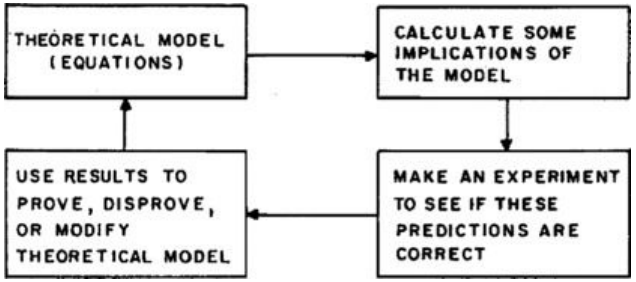


Cloud-flare

Ellen Airhart
WIRED
Jul 29, 2018

Mire jók a modellek

- Ha biztosítjuk, hogy jó leírásai legyenek a valóságnak, akkor
  - előrejelzik a számítógép működését
  - garantálható tulajdonságokat adnak
- De ez a viszony aluldeterminált

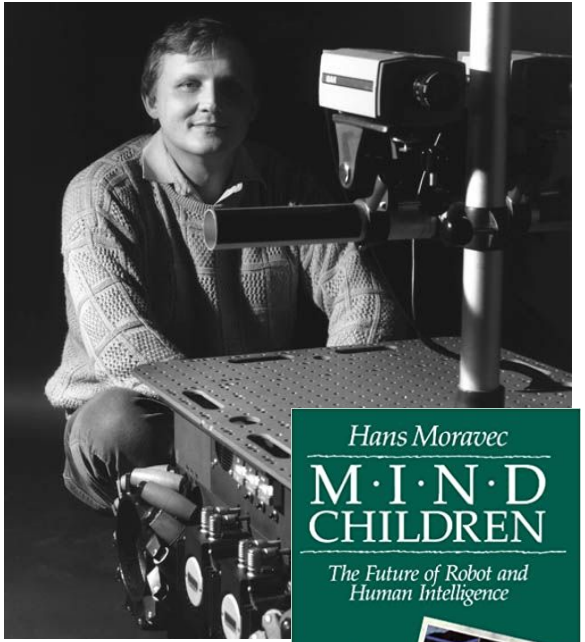


McCulloch

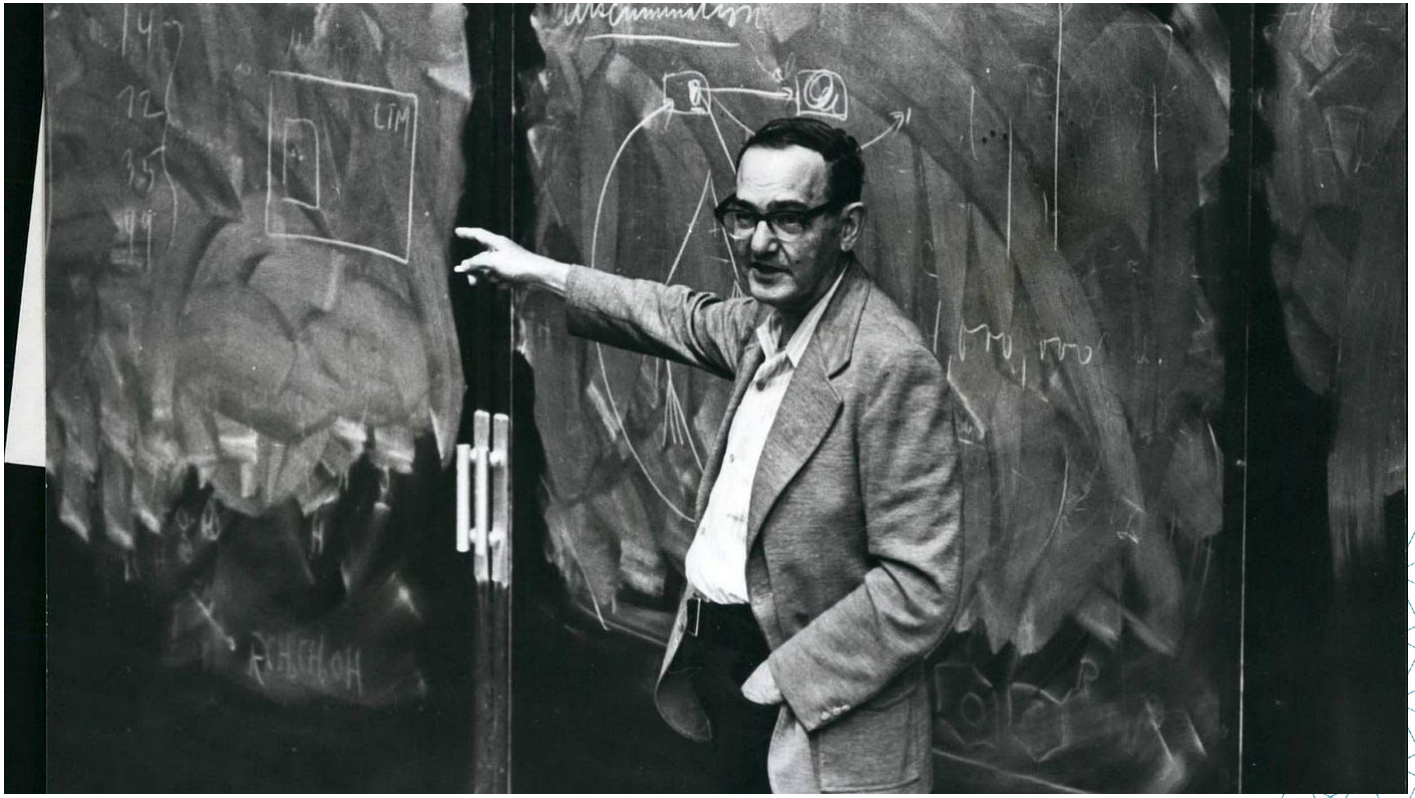
Wittgenstein

Antall

Agyprotézis gondolatkísérlet (ismétlés)



Herbert Simon 1969: the sciences of the artificial



Köszönöm szépen a figyelmet!

Héder Mihály

National Research, Development and Innovation Office
Hungary

PROGRAM
FINANCED FROM
THE NRDI FUND