



16 december 2020

Remissvar, Innovation genom information SOU 2020:55

Journalism++ Stockholm (J++) intresse i den här utredningen är dels som förädlare av öppen data när vi bygger nyhetstjänster och liknande, och dels som granskande journalister.

J++ ställer sig bakom det remissvar som undertecknats av en rad organisationer koordinerade av Civic Tech Sweden.

Utöver vad som framförs där, vill vi tillägga följande:

4 § 1. [... gäller inte för ...] datorprogram

Utredningen föreslår att den nya lagen inte ska gälla för information som utgör datorprogram. Här saknar vi en tydlig avgränsning av "datorprogram". Redan i dag finns en gråzon mellan vad som utgör källkod eller programkod, och vad som utgör *data* som hanteras av programkod. Den gråzonen växer.

Exempel 1: Maskinlärande datorprogram

Mycket av det som i dag kallas "AI" är maskininlärningsalgoritmer, där ett system tränas på, och självt förändras av, en datamängd. Internationellt är intresset för maskininlärning inom myndighetsutövning stort (se till exempel Algorithm Watch årsrapport 2020 för exempel på hur *predictive policing* nu kommit till Europa: <https://automatingsociety.algorithmwatch.org/report2020/>), och vi kan anta att det kommer att bli vanligt också i Sverige. I sådana system är gränsen mellan "datorprogram" och "data" svår att dra.

Samtidigt är träningsdatan som används för sådana modeller av stort intresse. Det är den som gör det möjligt för oss som journalister att granska algoritmbaserad myndighetsutövning (på detta finns otaliga exempel från USA), men det också data som har stor potential för vidareanvändning. Okontroversiell träningsdata för till exempel klassificering av dokument på svenska skulle ha stort värde för en mängd svenskspråkiga aktörer, under förutsättning att den kan definieras som just data, och inte kommer att betraktas som en del av ett datorprogram.

Exempel 2: Regelstyrda algoritmer

I våras begärde vi ut källkoden till en uppmärksammas "robot" som används av Trelleborgs kommun för att bedöma ansökningar om ekonomiskt bistånd (se

<https://www.dagenssamhalle.se/paket/lackan-fran-bistandsroboten-4109>). Frågan om källkoden skulle betraktas som allmän och offentlig handling hade då nyligen prövats av Kammarrätten. Det Trelleborgs kommun och kammarrätten båda kallade "källkod", "datorprogram" och "robot" visade sig dock snarast vara en regelsamling till ett proprietärt system/datorprogram för automatisering, och alltså knappast i sig själv vad man normalt skulle kalla ett "datorprogram". (Tekniskt sett kan man möjligen argumentera för att regelsamlingen utgör en beskrivning av en algoritm som därmed är att betrakta som programkod, men lika gärna för att detta var data som tagits fram av kommunen.) Begreppsförvirringen hos både kommunen och domstolen visar att redan med klassiska regelstyrda algoritmer som används av svenska myndigheter i dag, är gränsdragningen mellan kod och data svår. I fallet Trelleborg innehöll koden, eller datan, en mängd personuppgifter till personer som tidigare varit i kontakt med socialtjänsten i kommunen, som för att understryka det svåra i att dra en skarp gräns mellan kod och data.

För oss som granskande journalister är det förstås viktigt att kunna resonera med myndigheter om var gränsen går mellan programkod, som kanske omfattas av upphovsrätt eller affärssekretess, och data som skapas på myndigheten. Men detta är också data som har stort vidareförädlingsvärde. För en myndighet måste det vara tydligt vad det är som avgränsningen mot "datorprogram" utesluter.

Leonard Wallentin
Journalism++ Stockholm