Deutscher Bundestag **Drucksache 19/2844**

19. Wahlperiode 19.06.2018

Kleine Anfrage

der Abgeordneten Andrej Hunko, Dr. Alexander S. Neu, Heike Hänsel, Christine Buchholz, Anke Domscheit-Berg, Matthias Höhn, Ulla Jelpke, Zaklin Nastic, Tobias Pflüger, Eva-Maria Elisabeth Schreiber, Dr. Petra Sitte, Kathrin Vogler und der Fraktion DIE LINKE.

„Big Data“ und Software zur Vorhersage von „Krisen“ bei der Bundeswehr

Das Bundesministerium der Verteidigung (BMVg) plant die Beschaffung von Software zur Verarbeitung von großen Datenmengen, um damit „mögliche Ausstattungs- und Versorgungsprobleme“ zu identifizieren („Blick in die Zukunft: Big-Data-Software für die Bundeswehr“, bundeswehr-journal.de vom 2. Juni 2018). Hierzu setzt das BMVg auf die sogenannte In-Memory-Technologie, wonach die zu verarbeitenden Datenbestände im Arbeitsspeicher der Anwendung gehalten werden und damit um ein Vielfaches schneller durchsuchbar sind. Bereits im Jahr 2015 hatte das BMVg einen „Bedarf für den Einsatz einer In-Memory-Technologie“ mitgeteilt (Bundestagsdrucksache 18/5196, Antwort zu Frage 6). Damals wurden die Produkte „HANA“ des deutschen Konzerns SAP sowie „Blu“ von IBM für eine mögliche Beschaffung untersucht. Dem BMVg zufolge habe das Ministerium außerdem eine „grundsätzliche Marktanalyse/-sichtung“ zu Anwendungen zur Prognose von großen Datenmengen durchgeführt und mit einigen Firmen anschließend „ein Gespräch“ geführt (Bundestagsdrucksache 19/1979, Antwort auf die Schriftliche Frage 77 des Abgeordneten Omid Nouripour).

Derzeit testet das Bundesministerium „im Rahmen von Pilotprojekten“ die Prognosewerkzeuge SAP Analytics sowie „IBM Watson“. Das SAP-Programm diene demnach der „vorausschauende[n] Wartung (predictive maintenance)“. Mit dem IBM-Programm sollen außerdem „potenzielle Krisen“ vor ihrer Entstehung in den nächsten sechs bis 18 Monate vorhergesagt werden („Blick in die Zukunft: Big-Data-Software für die Bundeswehr“, bundeswehr-journal.de vom 2. Juni 2018). Hierfür muss die Software allerdings auf eine Datenbasis mit früheren politischen „Ereignissen“ zugreifen. Eine solche Datenbank „Global Data on Events, Location and Tone“ (GDELT) wird von dem Informatiker Kalev Leetaru betrieben (www.gdeltproject.org). Das Ziel von GDELT ist das Aufspüren zukünftiger Unruhen, Revolten oder Anschläge. GDELT wertet täglich Tausende Berichte aus öffentlich zugänglichen Medien aus. Auch das von Google und dem US-Geheimdienst CIA gegründete Unternehmen „Recorded Future“ bietet entsprechende Dienste an (www.recordedfuture.com). Unter anderem hatte das US-Militär derartige Verfahren 2011 im Rahmen seiner Intervention in Libyen eingesetzt (http://gleft.de/2hF). Hierfür werden öffentlich zugängliche Quellen im Internet ausgewertet (sogenannte OSINT-Verfahren). Zu einer solchen „IT-gestützten Nachrichtengewinnung aus offenen Quellen“ testete die Bundeswehr die beiden Analysetools „Textrapic“ und „Brandwatch“ zur Erfassung von „Meinungs- und Stimmungslagen der Bevölkerung“. Breiter aufgestellte Forschungen

für eine „Wissenserschließung aus offenen Quellen“ (WeroQ) hatte die Bundeswehr zwar für den Zeitraum von 2014 bis 2016 beauftragt, diese wurden aber wegen eines fehlenden Zuwendungsbescheides des Bundesamtes für Ausrüstung, Informationstechnik und Nutzung der Bundeswehr (BAAINBw) für das Fraunhofer-Institut zunächst nicht durchgeführt (Schreiben des Parlamentarischen Staatssekretärs Dr. Ole Schröder vom 22. Juli 2014, http://gleft.de/2hE) und fanden am Ende gar nicht statt (Plenarprotokoll 18/192). Auch das Bundeskriminalamt hatte unter Koordination des Bundes Deutscher Kriminalbeamter mit dem Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS die Software „FuhSen“ (föderierte und hybride Search-Engine) zur semantischen Ermittlung in offenen Quellen im Internet entwickelt, die neben gängigen Schnittstellen zu Social Media auch das „Deep Web“ (etwa Online-Marktplätze) oder das „Dark Web“ durchsucht (http://gleft.de/2hJ). Im Projekt wurde auch geprüft, wie eine polizeiliche Verarbeitung von Personendaten nach der EU-Datenschutz-Grundverordnung rechtskonform erfolgen kann.

Die jetzige militärische Beschaffung der Software SAP Analytics sowie IBM Watson ist Teil einer „Digitaloffensive“ der Bundeswehr. Vor der Entscheidung hatte das BMVg auch die von der hessischen Polizei genutzte Vorhersagesoftware „Gotham“ der US-Firma Palantir Technologies geprüft. Die Nutzung von SAP-Produkten kann unter anderem ein Sicherheitsrisiko darstellen. Nach Recherchen des ARD-Magazins „FAKT“ lieferte der deutsche Konzern seine Datenbanktechnologie „HANA“ an US-Geheimdienste (ARD FAKT vom 10. März 2015).

Wir fragen die Bundesregierung:

1. Aus welchen Gründen zeichnete sich im Jahr 2015 „Bedarf für den Einsatz einer In-Memory-Technologie ab“ (Bundestagsdrucksache 18/5196, Antwort zu Frage 6)?

   a) Welche Ergebnisse zeitigten die damaligen Untersuchungen der Produkte SAP SE (Produkt „HANA“) und IBM (Produkt „Blu“)?

   b) Auf welche Weise und von wem wurden die Produkte „untersucht“ bzw. erprobt?

2. Aus welchen Gründen wurde die geplante Studie „Wissenserschließung aus offenen Quellen“ (WeroQ) nicht durchgeführt (Plenarprotokoll 18/192, S. 19133 (D))?

   a) Welches Fraunhofer-Institut und welche weiteren Beteiligten sollten die Studie durchführen?

   b) Sofern der Zuwendungsbescheid durch das BAAINBw an das Fraunhofer-Institut niemals erfolgte, welche Gründe kann die Bundesregierung hierzu mitteilen (Schreiben des Parlamentarischen Staatssekretär Dr. Ole Schröder vom 22. Juli 2014, http://gleft.de/2hE)?

   c) Welche Forschungen zu „marktverfügbare[n] Analysetools“ zur Erfassung „von Meinungs- und Stimmungslagen der Bevölkerung“ hat die Bundeswehr in den letzten fünf Jahren betrieben bzw. beauftragt?

   d) Was ergab die Erprobung der beiden Analysetools „Textrapic“ und „Brandwatch“?

   e) Welche ähnlichen Forschungen zur „IT-gestützten Nachrichtengewinnung aus offenen Quellen“ (sogenannte OSINT-Verfahren) hat die Bundeswehr in den letzten fünf Jahren betrieben bzw. beauftragt (bitte auch mitteilen, welche Anwendungen „betrachtet“ wurden)?

3. Welche Soft-und Hardware welcher Hersteller hat das BMVg im Rahmen einer „grundsätzlichen Marktanalyse/-sichtung zu entsprechenden Produkten“ zur Prognose von großen Datenmengen geprüft, und aus welchen Erwägungen wurde sich gegen die Beschaffung entschieden (Bundestagsdrucksache 19/1979, Antwort auf die Schriftliche Frage 77 des Abgeordneten Omid Nouripour)?
4. Nach welchem Verfahren hat das BMVg eine solche Marktsichtung durchgeführt, bzw. wie wurden die zu sichtenden Produkte und Firmen gefunden?
5. Mit welchen Firmen hat das BMVg anschließend „ein Gespräch“ geführt, und wer nahm seitens der Bundeswehr und des Bundesministeriums daran teil?
6. Welche Soft- und Hardware welcher Hersteller hat das BMVg zur Prognose von großen Datenmengen beschafft, und welche weiteren Beschaffungen sind geplant („Blick in die Zukunft: Big-Data-Software für die Bundeswehr“, bundeswehr-journal.de vom 2. Juni 2018)?
   a) Welche Defizite sollen mit der Beschaffung ausgeglichen werden?
   b) Auf welche Weise soll die Technik helfen, „Ausstattungs- und Versorgungsprobleme bei der Bundeswehr“ zu erkennen und ihre Lösung zu befördern (bitte hierzu auch die technische Funktionsweise erläutern)?
7. Wann wurde die Software installiert, wann startete und wann endet der Wirkbetrieb im Pilotprojekt?
8. Inwiefern soll die Software „IBM Watson“ auch zur Vorhersage politischer Ereignisse („Krisen“) oder zur Bestimmung allgemeiner Meinungs- und Stimmungslagen genutzt werden?
   a) Welche Funktionalitäten der Software sind der Bundesregierung hierzu bekannt?
   b) Für welche Zeiträume verspricht die Software, dass „potenzielle Krisen“ vor ihrer Entstehung vorhergesagt werden könnten?
9. Welche Daten bzw. welche Quellen (auch OSINT) sollen mit der Software „IBM Watson“ verarbeitet werden (bitte für jedes Produkt einzeln darstellen)?
   a) Auf welche öffentlichen und nichtöffentlichen Datenbanken früherer Ereignisse (etwa „Global Data on Events, Location and Tone“) greift die Software zurück?
   b) Inwiefern und auf welche Art werden diese Quellen, deren eigene Quellen und ihre jeweilige Zuverlässigkeit evaluiert, bzw. inwieweit erfolgt eine Kontrolle und Validierung der gesammelten Daten durch andere Erkenntnisse (http://gleft.de/2hO)?
   c) Auf welche Regionen der Welt sind diese Quellen und Datenbanken bezogen?
   d) Werden die über diese Datenbanken erlangten Daten bei der Bundeswehr auch zu irgendeinem anderen Zweck genutzt (unabhängig von „IBM Watson“)?

10. Welche Konsequenzen ziehen das BMVg und die Bundeswehr für die geplante Krisenvorhersage mit „Watson“ aus den Erfahrungen mit „Tay“, dem auf künstlicher Intelligenz basierenden Chat-Bot von Microsoft, der im Jahr 2016 aufgrund der ihm im Netz zugänglich gewordenen Daten innerhalb kürzester Zeit begann, anderen Nutzern sexistische und rassistische Nachrichten zu schreiben (http://gleft.de/2i5; http://gleft.de/2i6), sowie den Erfahrungen mit IBMs „Watson“ selbst, das u. a. begann, sich in Obszönitäten auszudrücken, nachdem es im Netz den „Urban Dictionary“ gelesen hatte (http://gleft.de/2hP)?

11. Welche einmaligen Kosten entstanden für die Beschaffung von Software zur Vorhersage politischer Ereignisse („Krisen“), und welche jährlichen Lizenzen (auch für den Zugriff auf Datenbanken mit politischen Ereignissen) werden fällig, sollte sich das BMVg für einen Kauf entscheiden (bitte für jedes Produkt einzeln darstellen)?

    a) Welche Rolle spielt der „Cyber Innovation Hub“ der Bundeswehr bei der Beschaffung und dem Betrieb von digitalen Prognosewerkzeugen für das BMVg (http://gleft.de/2hy)?

    b) Welche privaten Firmen oder Institute gehören dem „Cyber Innovation Hub“ an?

12. Welche Einheiten der Bundeswehr sollen die Software zur Vorhersage politischer Ereignisse („Krisen“) nutzen?

    a) Welche militärischen Konsequenzen bzw. sicherheitspolitischen Erkenntnisse sollen ggf. aus Krisenmeldungen von „IBM Watson“ gezogen werden?

    b) Bis zu welcher Reaktionsstufe sollen im Falle einer von „IBM Watson“ gemeldeten „Krise“ seitens des BMVg oder der Bundeswehr Maßnahmen ergriffen werden können?

13. Welche weiteren Bundesbehörden (auch das Bundeskanzleramt) nutzen SAP HANA, planen eine Beschaffung oder haben das Produkt „untersucht“?

14. Welche konkreten Dienstleistungen hat das BMVg seit Beantwortung der Bundestagsdrucksache 18/5196 von SAP in Anspruch genommen, welche Produkte wurden beschafft, und welche Kosten entstanden dafür?

15. Wie viele SAP-Mitarbeiterinnen und -Mitarbeiter (auch von Tochterfirmen wie SAP NS2, Inxight, Sybase) sind derzeit in welchen Projekten für die Bundeswehr tätig?

16. Welche „Lehrgänge“ hat die US-Firma Attensity für die Bundeswehr durchgeführt, und welche vergleichbaren Veranstaltungen wurden von anderen US-Firmen organisiert (Bundestagsdrucksache 18/5196, Antwort zu Frage 7)?

17. Welche Erkenntnisse besitzt die Bundesregierung über Datenintegrität, Hintertüren, Black Boxes etc. der zu beschaffenden Software von SAP und IBM oder weiterer Anbieter bzw. über eine Zusammenarbeit der Hersteller mit Geheimdiensten, bzw. inwiefern wurde dies überhaupt geprüft?

18. Werden nur das BMVg und die Bundeswehr mit Programmen ausgestattet und mit Programmen arbeiten, die (mit oder ohne In-Memory-Technologie) „potenzielle Krisen“ oder Betriebsabläufe vorhersehen sollen, oder auch beispielsweise das Auswärtige Amt, das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, das Bundesministerium des Innern, für Bau und Heimat (BMI) oder die Geheimdienste?

19. In welchen Vorhaben forschen Behörden des BMVg (auch die Universitäten der Bundeswehr) oder des BMI derzeit zu den Feldern

    a) Verbesserung von automatisierten Verfahren des „Data Minings“,

    b) Verarbeitung von „Massendaten“ in (nahezu) Echtzeit,

    c) In-Memory-Technologie,

    d) „Prediktive Analyse“ bzw. „Vorhersagende Schlussfolgerungen“ oder Ausgabe von kriminalistischen „Hypothesen“,

    e) Computergestützte Auswertung von sozialen Medien (darunter Twitter, Facebook),

    f) Analyse sozialer Netzwerke oder Suchmaschinen zum Aufspüren von Gefahren,

    g) Visuelle Darstellung öffentlich zugänglicher Open-Source-Informationen als automatisiertes datenbankgestütztes Tool zur Datensammlung, Auswertung, Analyse,

    h) Intelligente Auswertung von Sensoren im öffentlichen Raum (Bundestagsdrucksache 18/7966)?

20. Welche Erfahrungen hat das Bundeskriminalamt als assoziierter Partner aus dem im Rahmen der Förderbekanntmachung zum Themenfeld „Zivile Sicherheit – Schutz vor organisierter Kriminalität“ im Rahmen des Programms „Forschung für die zivile Sicherheit II“ eingereichten Forschungsprojekt Linked-Data-basierte Kriminalanalyse (LiDaKrA) gewonnen (Bundestagsdrucksache 18/7966, Antwort zu Frage 7)?

    a) Wer nahm an dem Projekt teil, und welche Produkte oder Verfahren wurden dabei entwickelt?

    b) Die Zusammenführung welcher polizeilicher, öffentlicher oder sonstiger Datenquellen (auch „Dark Web“) wurde in dem Projekt beforscht?

    c) Welche „von den Anbietern angebotenen Schnittstellen zu öffentlichen Quellen“ wurden dabei genutzt (Bundestagsdrucksache 18/8323, Antwort zu Frage 11)?

Berlin, den 13. Juni 2018

Dr. Sahra Wagenknecht, Dr. Dietmar Bartsch und Fraktion

Satz: Satzweiss.com Print, Web, Software GmbH, Mainzer Straße 116, 66121 Saarbrücken, www.satzweiss.com
Druck: Printsystem GmbH, Schafwäsche 1-3, 71296 Heimsheim, www.printsystem.de
Vertrieb: Bundesanzeiger Verlag GmbH, Postfach 10 05 34, 50445 Köln, Telefon (02 21) 97 66 83 40, Fax (02 21) 97 66 83 44, www.betrifft-gesetze.de
ISSN 0722-8333