

Bachberger-Strolz, Kerstin

Article

Profiling, Targeting, Algorithmen, künstliche Intelligenz – über die Irrwege einer Debatte in der Arbeitsmarktpolitik

Wirtschaft und Gesellschaft (WuG)

Provided in Cooperation with:

Kammer für Arbeiter und Angestellte für Wien

Suggested Citation: Bachberger-Strolz, Kerstin (2020) : Profiling, Targeting, Algorithmen, künstliche Intelligenz – über die Irrwege einer Debatte in der Arbeitsmarktpolitik, Wirtschaft und Gesellschaft (WuG), ISSN 0378-5130, Kammer für Arbeiter und Angestellte für Wien, Wien, Vol. 46, Iss. 3, pp. 329-363,
<https://doi.org/10.59288/wug463.2>

This Version is available at:

<https://hdl.handle.net/10419/333063>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0>

Profiling, Targeting, Algorithmen, künstliche Intelligenz – über die Irrwege einer Debatte in der Arbeitsmarktpolitik

Kerstin Bachberger-Strolz

1. Einleitung

„Alles ein großer Pallawatsch“, möchte man feststellen angesichts der bisherigen Diskussionen zum sogenannten „AMS-Algorithmus“! Allem Anschein nach fällt es den Beteiligten in Österreich schwer, eine öffentliche Debatte einigermaßen vernünftig und sachlich zu führen. Den Ausgang nahmen die Überlegungen zum „AMS-Chancen-Modell“ (auch AMAS, Arbeitsmarkt-Assistenzsystem, genannt) etwa im Jahr 2015, als mit steigender Arbeitslosigkeit und mit nicht im selben Ausmaß steigenden budgetären Mitteln in der aktiven Arbeitsmarktpolitik darüber debattiert wurde, ob und wie die Ausgaben zielgerichteter eingesetzt werden könnten. Vielleicht sollten manche Gruppen an Arbeitslosen auch weniger oder überhaupt nicht mehr gefördert werden, so die mehr oder minder explizit formulierte Fragestellung.

Es hob jedenfalls ein Stimmengewirr an, in dem eine Vielzahl von Begriffen wie „Segmentierung“, „Effektivität und Effizienz“, „Sozialpolitische Verantwortung“, „Profiling“, „Algorithmus“, „künstliche Intelligenz“ etc. kunterbunt durcheinandergewirbelt wurden. Besonders symptomatisch war die Verwechslung der Begriffe *Profiling*¹ und Algorithmus.² Dies führte zu zwei folgenschweren Irrtümern, von denen sich die Diskussionen bis dato nicht „erholt“ haben:

(1) Auf einer „technischen Ebene“ ging man nämlich davon aus, dass die Wiederbeschäftigungswahrscheinlichkeit eine gute Proxy-Größe für die Zuweisung einzelner Personen zu verschiedenen Kursen und Programmen wäre. Das ist sie aber eben gerade nicht! Eine Erkenntnis, die einfach durch das Studium der internationalen Erfahrungen zu gewinnen gewesen wäre.³

(2) Ein weiteres Missverständnis auf einer eher allgemein-politischen Ebene gab es in Kommentaren, die von außerhalb des arbeitsmarktpolitischen Diskurses kamen: Es wurde nämlich unterstellt, dass die Feststellung eines empirischen Sachverhaltes (nämlich die unterschiedlichen Arbeitsmarktchancen von Arbeitslosen) zu einem diskriminierenden Einsatz

der aktiven Mittel des AMS führen würden. Das ist natürlich nicht notwendigerweise der Fall!

Für einen Teil dieser Missverständnisse und Irrtümer ist jedoch das AMS selbst verantwortlich zu machen, weil es von Anfang an nicht klargelegt hat, was mit den *Profiling*-Ergebnissen letztlich gemacht werden wird. In der Realität wird sich wahrscheinlich sehr wenig ändern, weil die Zielgruppenlogik bei den Programmen nicht geändert wird. Damit stellt sich aber die Frage: Wozu dann überhaupt dieser große Aufwand? Bisher überhaupt außer Acht gelassen wurde darüber hinaus die Tatsache, dass in anderen Ländern *Profiling* als ein Frühwarnsystem im Falle von schlechten Beschäftigungsaussichten („Hier besteht besonderer Handlungsbedarf!“) verwendet wurde und wird. In diesem Falle stehen die Personen mit schlechten Beschäftigungsaussichten im Mittelpunkt der Aufmerksamkeit.

Insgesamt sind bei der Einführung des *Profiling*-Systems bisher einige folgenschwere Fehler gemacht worden: Es mangelt an Transparenz (eine Grundforderung für wissenschaftliches Arbeiten), wie die einzelnen Ergebnisse konkret zustande gekommen sind. Es fehlt die sozialwissenschaftliche Begleitung, die sich insbesondere der ethischen Fragen widmen müsste. Es fehlt darüber hinaus die Einbindung der Betroffenen, also sowohl der Arbeitslosen als auch der AMS-BeraterInnen. Und es fehlt eine klare Festlegung, was mit dem *Profiling* ganz konkret gemacht wird.⁴ Darüber hinaus wissen wir durch einen Spruch der Datenschutzbehörde vom 16. August 2020, dass für das *Profiling* eine entsprechende gesetzliche Grundlage und eine Datenschutz-Folgeabschätzung fehlen. Die Zukunft des *Profiling* im AMS steht damit in den Sternen. All diese Versäumnisse lassen sich durch private Blog-Beiträge von AMS-Vorständen jedenfalls nicht aus der Welt schaffen, insbesondere dann nicht, wenn diese Beiträge über eine affirmative, beschönigende und propagandahafte Beschreibung der Sachverhalte nicht hinausreichen.

Im Folgenden wird auf die Möglichkeiten und Grenzen von *Profiling* (Kapitel 2), auf die Frage „*Profiling* vs. *Targeting*“ (Kapitel 3) und auf die Probleme bei der Anwendung von künstlicher Intelligenz in der Sozial- und Arbeitsmarktpolitik (Kapitel 4) eingegangen.

2. Wofür eignet sich *Profiling* von Arbeitslosen, wofür nicht?

Die auf internationaler Ebene geführten Diskussionen zur Anwendung des statistischen *Profiling* in der Arbeitsmarktpolitik reichen zumindest bis in die 1990er-Jahre zurück.

2.1 Die Anfänge der Diskussionen⁵

Für das frühe Verständnis von statistischem *Profiling* im angloamerikanischen Raum bezeichnend ist der Titel einer OECD-Publikation aus dem Jahr 1998: „Early Identification of Jobseekers at Risk of Long-term Unemployment. The Role of Profiling“,⁶ die Ergebnis einer ExpertInnenkonferenz mit VertreterInnen aus den USA, CAN, AUS und UK war. Es ging also v. a. darum, Personen mit einer hohen Wahrscheinlichkeit, langzeitarbeitslos zu werden, frühzeitig zu identifizieren und die entsprechenden Schritte und Maßnahmen einzuleiten.

„The early identification of jobseekers at risk of becoming long-term unemployed is a longstanding and basic endeavour of the public employment services (PES). Indeed, good judgement in this area forms part of the professional competence and work experience of PES staff. However, a few countries have gone further by introducing more formal methods of identifying at-risk jobseekers and laying out procedures on what to do with them. This is usually referred to as profiling and is used in this paper to cover the approach of i) the identification of individuals at risk of long-term unemployment; ii) the referral to various active labour market programmes.“⁷

Im Jahr 2004 hat man in Schweden im Rahmen einer internationalen Konferenz („Can statistical models increase the efficiency at the employment office?“, Uppsala) überlegt, statistisches *Profiling* in der öffentlichen Arbeitsmarktverwaltung verstärkt einzusetzen. Der Direktor des IFAU (Institut für Arbeitsmarktpolitik-Evaluierung), Erik Mellander, fasste zu Beginn der Konferenz die zentralen Herausforderungen in Bezug auf *Profiling* zusammen: 1) Diese Methode soll als Entscheidungshilfe dienen und eine bessere Selektion ermöglichen, durch die effiziente Verwendung der vorhandenen Informationen; 2) die Effizienz erhöhen, d. h. die Nettoerträge (Erträge minus Aufwendungen) der Maßnahmen maximieren, und 3) die Chancengleichheit erhöhen. Die gesamte *Profiling*-Debatte wurde also schon zu Beginn überlagert von einer noch fundamentaleren Fragestellung, nämlich jener, ob die Nettoerträge maximiert werden (Effizienzkriterium) oder die bedürftigsten Personen (Gleichheitsziel) gefördert werden sollen. Nur wenn die größten Wirkungen bei den Personengruppen auftreten, die die höchste Wahrscheinlichkeit aufweisen, langzeitarbeitslos zu werden, fallen die beiden Zielsetzungen zusammen. Technisch gesprochen können also bei homogenen Maßnahmenwirkungen und homogenen Grenzkosten grundsätzlich keine Effizienzvorteile lukriert werden. In Berger et al. (2001, S. 65) ist der grundlegende Sachverhalt folgendermaßen formuliert: „... the optimal profiling variable is the one that maximises the attainment of the goals of the allocation mechanism“. Die beiden Fragen nach der Prognose der Langzeitarbeitslosigkeits-Wahrscheinlichkeit und nach den Maßnahmenwirkungen sind also grundsätzlich voneinander

verschieden, wenn auch alle *Profiling*-Anstrengungen darauf abzielen, die „richtigen“ Personen in die richtigen Maßnahmen zu bekommen – also jene Personen, bei denen sich die Bezugsdauer von Arbeitslosengeld am stärksten durch die Maßnahmenteilnahme verkürzt und/oder bei denen der Maßnahmeneffekt (etwa ein hoher Beschäftigungsanteil nach Teilnahme) am größten ist.

Berger et al. (2001, S. 59) fassen ihre Ergebnisse wie folgt zusammen:

„First, we demonstrate that it is possible to effectively predict the duration of UI spells⁸ ... This finding is important because effective prediction of the profiling variable is a necessary but not a sufficient condition for the success of a profiling system. Second, we show that the impact of reemployment service⁹ does not appear to vary with the expected duration of the UI spell, indicating that the UI profiling in Kentucky does not advance the goal of efficiency, though it may advance equity goals.“

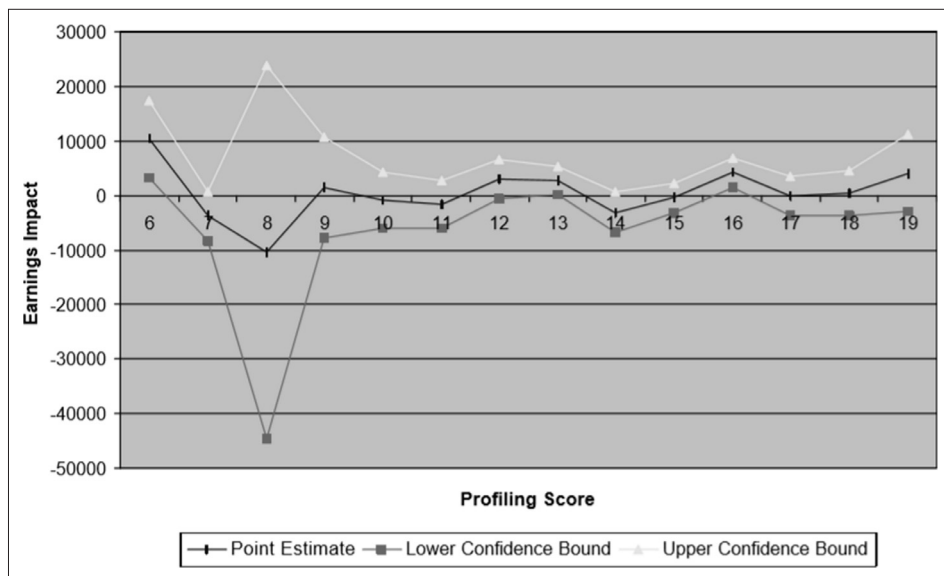
In einer weiteren Publikation kommen Black et al. (2003a, S. 1325) zu folgenden Schlussfolgerungen:

„Finally, the underlying assumption of the WPRS (Worker Profiling and Reemployment Service) program is that those with the longest expected UI spell durations would benefit the most from the requirement that they participate in reemployment services in order to continue receiving their UI benefits. It is also assumed that treating these claimants will result in the largest budgetary savings for state UI systems. Our results provide little justification for either assumption, as we do not find a monotone relationship between the profiling score and the impact of treatment. If the goal of profiling is to allocate the treatment to those claimants with the largest expected impact from it, or to save the state UI system the most money, then our findings call into question the wisdom of using the expected benefit duration as a means of allocating treatment.“¹⁰

Abbildung 1 veranschaulicht den – in den vorangegangenen Zitaten erwähnten – fehlenden Zusammenhang noch einmal grafisch: Zwischen der Länge der prognostizierten Dauer der Arbeitslosigkeit (*Profiling*-Score) und dem Ergebnis des „*Worker Profiling and Reemployment Service*“ (WPRS) (hier der Einkommenseffekt) besteht kein (!) Zusammenhang. *Profiling* anhand der Reintegrationschancen in den Arbeitsmarkt hilft also nicht, die Effizienz von Maßnahmen (WPRS) zu erhöhen. Entlang der prognostizierten Arbeitslosigkeitsdauer gibt es keine Maßnahmenheterogenität, das WPRS-Programm wirkt also für voraussichtlich langzeitarbeitslose Personen ähnlich wie für voraussichtlich kurzzeitarbeitslose.

Neben diesen grundsätzlichen Fragestellungen wurden frühzeitig in der Debatte zum statistischen *Profiling* auch technische Aspekte der Modellbildung erörtert, die auch heute noch relevant sind: Wie genau können die Profile erstellt werden? Welche Anforderungen an die Daten müssen beim *Profiling* erfüllt sein? In einer Machbarkeitsstudie für Großbritannien ver-

Abbildung 1: Wirkungen auf die Einkommen entlang der *Profiling-Scores*



Quelle: Berger et al. (2001) 81.

suchen Bryson und Kasparova (2003) nicht, die beste *Profiling*-Methode zu ermitteln, sondern mit der Untersuchung sollte die Sensitivität der Ergebnisse hinsichtlich einzelner Modellspezifikationen und anderer Faktoren illustriert werden.¹¹ Es wurden drei Ergebnis-Variablen¹² für die drei Hauptgruppen von LeistungsbezieherInnen „Kranke/Behinderte“, „Allein erziehende Eltern“ und „AIG-BezieherInnen“ prognostiziert. Zwölf Pilotgebieten wurden zwölf zufällig ausgewählten Vergleichsregionen gegenübergestellt. Die zentrale Frage der ausreichenden Genauigkeit der Prognosen wurde über zwei Verfahrensschritte angegangen: a) den Vergleich der Durchschnitte einzelner Quintile im Vergleichssample und b) durch die Festlegung einer Entscheidungsregel (30%, 50% oder 70% Maßnahmen-teilnehmer) und die anschließende Ermittlung von „korrekter Teilnahme“ (das Modell prognostiziert im Vergleichssample – *out of sample prediction* – eine Maßnahmenteilnahme für eine Person, die tatsächlich an einer Maßnahme teilgenommen hat), „korrekter Nichtteilnahme“, „nicht korrekter Teilnahme“ und „nicht korrekter Nichtteilnahme“. Eine zentrale Annahme in dieser Machbarkeitsstudie besteht darin, davon auszugehen, dass Personen mit der höchsten Wahrscheinlichkeit, langzeitarbeitslos zu werden, die höchsten Maßnahmenwirkungen erzielen.

Die Ergebnisse der Untersuchung sind folgende:

- *Profiling* ist besser als eine Zufallszuteilung zu Maßnahmen – eine vergleichsweise niedrige Hürde! „*Wrong denial*“ und „*wrong treat-*

ment“ bleiben jedoch ein gravierendes Problem (im Ausmaß von 26 bis 44% aller Personen)!

- Die Genauigkeit der Prognose hängt wie bei vielen statistischen Verfahren von der Varianz der *Outcome*-Variable (höher ist besser) ab; zudem vom Anteil der Personen, die in irgendeine Maßnahme kommen; und von der Verfügbarkeit von erklärenden Variablen.
- Die verschiedenen Schätzmethoden zeigen keinen großen *Performance*-Unterschiede (dieses Ergebnis wird von anderen Studien nicht bestätigt; siehe unten).
- Modelle mit mehr erklärenden Variablen prognostizieren meist etwas genauer.
- Wenn die Gruppe der Anspruchsberechtigten groß ist (70% *treated*), dann können etwa 75% der Personen richtig vorhergesagt werden. Wenn die Maßnahmenteilnahme jedoch stärker fokussiert angeboten wird, dann fällt die Prognosegüte auf etwa 58%; in diesem Fall fällt auch die Anzahl der „*wrongly treated*“, aber die Anzahl der fälschlicherweise Abgelehnten steigt deutlich an. Dieses Ergebnis würde gegen ein starkes „*Targeting*“ sprechen.

Black et al. (2003b) gehen ebenfalls Fragen der Modellentwicklung im Rahmen des *Profiling* nach, mit folgenden Ergebnissen:

- Kleinstquadratschätzer mit stetiger, abhängiger Variablen sind geeigneter als nicht-lineare Logit- und Tobit-Modelle.
- Werden als abhängige Variable stetige Größen (wie die Anteile der Personen, die ihren ALG-Anspruch verbraucht haben) verwendet, sind bessere Ergebnisse zu erzielen als bei binären Variablen (null/eins oder ja/nein).
- Modelle mit mehr erklärenden Variablen verbessern die Qualität.
- Die Modelle sollten im Verlauf der Zeit neu geschätzt werden, weil ihre *Performance* über den Konjunkturzyklus variiert.

Daraus folgt, dass *Profiling* mit erheblichen Unsicherheiten in der Modellbildung verbunden ist. Das schwerwiegendste Problem dabei ist die Schätzungenauigkeit, die dazu führt, dass Personen in die falschen Gruppen zugeteilt werden: Arbeitslose mit einer niedrigen Wahrscheinlichkeit, langzeitarbeitslos zu werden, werden als Personen mit hoher Langzeitarbeitslosigkeits-Wahrscheinlichkeit eingestuft und umgekehrt. Wenn das auch nur für 20-30% der Gesamtpopulation geschieht, ist das eine erschreckend hohe Zahl, die in Österreich etwa 180.000-270.000 Personen betreffen könnte (siehe Kapitel 2.4).

2.2 Die Diskussion in den 2010er-Jahren¹³

Nach der ersten Euphorie über die Einsatzmöglichkeiten von *Profiling* in der Arbeitsmarktpolitik ebte die Diskussion stark ab, um erst in den

2010er-Jahren wieder fortgesetzt zu werden. Gegenwärtig verwenden elf Länder in der OECD statistisches *Profiling*, was aber im Umkehrschluss heißt, dass 25 Staaten die Anwendung dieses Instruments nicht für sinnvoll erachten. Welche Erfahrungen wurden mit *Profiling* in etwa 25 Jahren des Einsatzes gemacht, wofür eignet sich *Profiling*, und wofür eignet es sich nicht? Die OECD kommt zu folgendem Schluss:

„Accurate profiling tools improve cost-efficiency of PES by reducing dead-weight costs, i.e. the cost related to providing services to jobseekers who would have found a job in any case, and by targeting resources to jobseekers most in need of help ... Many of the key issues raised in 1998 are still relevant today such as the choice of a profiling approach, the accuracy of profiling tools, the link with service delivery and ensuring buy-in by case workers.“¹⁴

Statistisches Profiling wird also gegenwärtig verwendet, um Zielgruppen mit niedrigen, mittleren und höheren Integrationschancen zu definieren; um den Zeitpunkt und die Intensität der Betreuung und die Zuweisung zu Programmen zu verbessern; um die Arbeit der Berater zu ergänzen (beispielsweise ist in Australien, Irland und den USA die Verwendung von Prognose-Modellen verpflichtend; die Ergebnisse legen die Leistungen fest) und sie zu unterstützen (freiwillige Nutzung beispielsweise in Dänemark und Schweden; Unterstützung der Entscheidungsfindung beispielsweise in Belgien); und um die Kosteneffizienz zu steigern. Neu hingegen ist die Verwendung von Verhaltensdaten, also Anklickdaten oder Befragungen, um Aspekte wie Suchintensität und Motivation abbilden zu können. Und neu ist auch der Einsatz von Künstlicher Intelligenz bzw. elaborierten Maschinenlern-Techniken (etwa in Flandern/Belgien, Dänemark, Neuseeland).¹⁵

Welche Probleme mit statistischem *Profiling* konnten in den letzten 25 Jahren nicht gelöst werden?

- Nach wie vor offen bleibt die Frage nach der Wahl des „richtigen“ *Profiling*-Zuganges bzw. wie die Genauigkeit der Schätzergebnisse, die häufig im Bereich von 60-80% liegen, verbessert werden kann. Die Präzision der Schätzungen hängt offensichtlich von der Qualität und Art der *Input*-Daten ab, aber nicht von der Wahl des Modells (hier ist die Evidenz nicht eindeutig) und auch nicht von der Verwendung von Verhaltensdaten.¹⁶ Aber die Legitimität und die Zweckmäßigkeit (der Nutzen) der Verwendung von *Profiling* steht in einem unmittelbaren Zusammenhang mit der Genauigkeit der Ergebnisse – wenn jedoch 20-40% der Personen falsch zugeordnet werden („*wrongly treated*“ und „*wrongly denied*“), dann wird die Brauchbarkeit diese Instruments fundamental in Frage gestellt.¹⁷ Dabei handelt es sich um eine besonders ausgeprägte Form von statistischer Diskriminierung, die ja immer dann auftritt, wenn Personen mit bestimmten Merkmalen, die vom

Durchschnitt aus verschiedenen Gründen abweichen, dennoch an diesem Durchschnitt gemessen werden.¹⁸

Langenbucher (2019) nennt noch zwei ebenfalls wichtige Punkte:

- **Verzerrte Daten:** „Automatisierte Urteile sind so gut wie die Daten, mit denen sie programmiert werden; Daten sind repräsentativ für die Vergangenheit, aber nicht notwendigerweise für Gegenwart und Zukunft; systematische Fehler können unbemerkt bleiben und sich über die Zeit verstärken.“
- **Mangel an Transparenz:** „der Algorithmus ist oft schwierig zu verstehen; er kann nicht überprüft werden (v. a. Problem bei Modellen, die mit künstlicher Intelligenz arbeiten)“.

Weitere ungelöste Probleme sind:

- Es gibt offensichtlich einen erheblichen institutionellen Widerstand auf Seiten der BeraterInnen. Scott Gibbons, ein Vertreter aus dem US-Department of Labour, fasste den Sachverhalt 2018 folgendermaßen zusammen:

„Worker Profiling is well established in the US, but not well accepted; ironically, the largest challenge we experience is developing buy-in with front line staff (low receptivity to the concept of the program – don’t want to be told who to serve, how to serve; low willingness to serve people involuntarily; very low willingness to take adversarial action).“¹⁹

Interessant dabei ist v. a. die Tatsache, dass diese Aussage nach 25 Jahren (!) Erfahrungen in den USA getroffen wird!

- Ebenfalls weitgehend ungelöst ist die Frage, was mit den drei Gruppen mit niedrigen, mittleren und höheren Integrationschancen genau gemacht werden soll: Wenn Mitnahmeeffekte gering gehalten werden sollen, dann ist klar, dass mit zur Beschäftigungsaufnahme fähigen Personen wenig Aufwand seitens des PES getrieben werden sollte. Während allerdings im Großteil der Länder der Fokus auf Arbeitslose mit niedrigen Integrationschancen gelegt wird, steht in den USA und in Österreich die mittlere Gruppe im Zentrum der Reintegrationsbemühungen.²⁰
- Schließlich wird im obigen Zitat von Desiere et al. (2019) etwas verstoßen („*the link with service delivery*“) auf das zentrale Problem beim *Profiling* hingewiesen: Es wurde auch angesichts der Erfahrungen der letzten 25 Jahre noch immer nicht hinreichend klar, was mit den Ergebnissen von *Profiling* genau gemacht werden soll bzw. wie diese verwendet werden können, um die Ergebnisse der Programmteilnahmen zu verbessern.

Auf die Frage, was man angesichts der offensichtlichen Grenzen des Einsatzes von *Profiling* tun könnte, stellt Langenbucher (2019) fest, dass die Daten regelmäßig aktualisiert werden müssten; umfangreichere Datensätze eingesetzt werden sollten; dass die Ergebnisse der Arbeit von

BeraterInnen, Algorithmen und Kombinationen von beiden regelmäßig evaluiert werden müssten; und dass alle Akteure eingebunden werden sollten, d. h. insbesondere BeraterInnen und Arbeitslose.

2.3 Zusammenfassung der internationalen Erfahrungen mit *Profiling*

Wie bereits erwähnt, fällt es auf, dass nach mehr als 25 Jahren Erfahrungen zwei Drittel der OECD Länder statistisches *Profiling* nicht (!) nützen; manche Länder wie etwa Deutschland und die Schweiz haben (Pilot-)Versuche mit *Profiling* gestartet, sind aber wieder davon abgekommen.

Welche Schlüsse sollten daraus gezogen werden? Es ist offensichtlich auch nach beinahe drei Jahrzehnten noch immer nicht in einem befriedigenden Ausmaß gelungen, die mit *Profiling* in Zusammenhang stehenden Probleme so zu lösen, dass die mit diesem Instrument verbundenen Vorteile die Kosten der Datenaufbereitung, -wartung, -aktualisierung, der Schulung von BeraterInnen etc. eindeutig überragen. Diese Erkenntnis stellt den internationalen Erfahrungsrahmen für die Überlegungen zur Einführung von *Profiling* in Österreich dar.

2.4 Die Umsetzung von *Profiling* im AMS: eine *Blackbox*!

Abgesehen von Vorläufern²¹ hat die aktuelle Debatte zum *Profiling* im AMS etwa im Jahr 2016 begonnen.²² Die Synthesis-Forschung Ges.m.b.H. wurde vom AMS beauftragt, ein „Integrationschancen-Prognosemodell zur Kundensegmentierung im AMS“ zu entwickeln. In einer sehr knappen Darstellung auf nicht einmal zehn spärlich beschriebenen Seiten wurde die methodische Vorgangsweise – für die breitere Öffentlichkeit zugänglich – dokumentiert.²³ Demnach hat Synthesis ein „Modell zur Prognose der regionalspezifischen Arbeitsmarkt-Integrationschancen von vorgemerkten Arbeitslosen“²⁴ mittels logistischer Regressionen geschätzt, wobei das Modell persönliche Merkmale (Geschlecht, Alter, Staatsbürgerschaft, Ausbildung, Betreuungspflichten, gesundheitliche Einschränkungen), den bisherigen Erwerbsverlauf und den aktuellen Geschäftsfall (hier werden bisheriger Beruf, Ausmaß der Beschäftigung, Häufigkeit und Dauer von AMS-Geschäftsfällen, etwaiger Maßnahmeneinsatz und Typ des regionalen Arbeitsmarktgeschehen) berücksichtigt. Von ca. 1,2 Mio. AMS-Geschäftsfällen im Jahr konnten im Basismodell nur für ca. 440.000 Fälle „voll valide Schätzungen“²⁵ durchgeführt werden, weil nur für diese Personen im Zeitraum von vier Jahren vor Arbeitslosigkeit lückenlose Informationen zur Verfügung standen (dies ist eine wichtige Information für die Beurteilung der Treffergenauigkeit bzw. Präzision der Schätzungen).

Als Zielkriterien, die die „gelungene Integration“ darstellen sollen, wurde (1) in der kurzfristigen Perspektive angenommen, dass im Zeitraum von 7 Monaten 3 Monate in ungeförderter Beschäftigung verbracht wurden;

und (2) wurde in der langfristigen Perspektive von 6 Beschäftigungsmonaten im Zeitraum von 24 Monaten ausgegangen. Die Segmentierung der Arbeitslosen wurde schließlich – „unter Berücksichtigung der Trefferquote“, wie es in der Darstellung heißt – folgendermaßen vorgenommen: Die Gruppe mit hohen Integrationschancen ergibt sich aus den Personen, bei denen der Wert der kurzfristigen Integrationschancen bei über 66% liegt; bei jenen mit niedrigen Chancen liegt die Wahrscheinlichkeit, das langfristige Kriterium zu erreichen, bei weniger als 25%. Die restlichen Personen bilden die mittlere Gruppe ab.²⁶

In einem auch öffentlich bekannt gewordenen Schätzergebnis²⁷ werden die kurzfristigen Integrationschancen der Basispopulation gezeigt: Bezogen auf eine Referenzgruppe,²⁸ für die die kurzfristige Integrationschance bei 52% liegt (die Konstante von 0,10 wurde logistisch transformiert), zeigt sich etwa, dass die Merkmale Geschlecht, Alter 30-49 und über 50 Jahre, aus einem Drittstaat, etc. die Integrationschancen verringern; eine Ausbildung über dem Pflichtschulniveau, aus der EU kommend oder im Produktionssektor vorher beschäftigt, etc. verbessern hingegen die Chancen, kurzfristig eine Beschäftigung zu finden. Bleiben alle anderen Merkmale gleich und es ändert sich beispielsweise nur das Geschlecht, dann ist ein Wert von 0,14 abzuziehen, sodass sich die Integrationswahrscheinlichkeit geringfügig auf 49% vermindert. Damit wird ein empirisches Faktum festgestellt, das die Diskriminierung von Frauen am Arbeitsmarkt zeigt. Ob das AMS daraus „diskriminierende Schlüsse“ zieht, wäre eine ganz andere Frage. Das Kriterium „Altersgruppe 50+“ wiederum reduziert die Wahrscheinlichkeit auf kurzfristige Wiedereingliederung sogar auf 35% (-0,70). Mit diesem Schätzergebnis lässt sich allerdings nicht sinnvoll erklären, warum etwa ein Universitätsabschluss keinen Einfluss auf die Integrationschancen haben sollte oder warum eine höhere Frequenz an Geschäftsfällen im Zeitraum von vier Jahren die Wiederbeschäftigungswahrscheinlichkeit derart nach oben treiben sollte (+1,98). Die letzte Variable hat offensichtlich das falsche Vorzeichen, was von Synthesis unkommentiert bleibt.

Vergleicht man die Ergebnisse von Synthesis über verschiedene Jahre, dann lassen sich Rückschlüsse auf die Stabilität der Schätzergebnisse in Bezug auf die Zuordnung zu den Gruppen mit niedrigen („C-Bereich“) und mit hohen („A-Bereich“) Reintegrationschancen ziehen. Die Verteilung für die valide schätzbaren Personengruppen zeigt, dass 33% zur Gruppe mit niedrigen und 7% zur Gruppe mit hohen Integrationschancen zählen. Das ist insofern interessant, als die Verteilung 3 Monate zuvor noch 4% zu 32% lautete.²⁹ Ob diese drastische Änderung eines wichtigen Hauptergebnisses auf methodische Änderungen zurückzuführen ist, wird aus den Unterlagen nicht deutlich. Darüber hinaus interessant sind die großen regionalen Unterschiede in der Verteilung der Arbeitsmarktgruppen, insbesondere zwischen Wien und Tirol.

2.4.1 Kritik am AMS-Profiling-Ansatz

Die Kritik daran, wie das AMS *Profiling* bisher implementiert hat, ist vielfältig: Sie reicht vom Vorwurf der intransparenten Vorgangsweise, der Diskriminierung einzelner Gruppen, der Anwendung von Künstlicher Intelligenz ohne sozialwissenschaftliche Begleitung, der fehlenden Einbindung von Betroffenen (BeraterInnen, Arbeitslose), des mangelhaften Datenschutzes bis zur der Verletzung von Grundrechten durch die Verwendung von Algorithmen u. a. Nachfolgend wird v. a. auf technische Aspekte eingegangen.

Mangelnde Transparenz: Auch wenn AMS-Vorstand Kopf nicht müde wird, das Gegenteil zu behaupten:³⁰ Das AMS-*Profiling* wurde bisher völlig intransparent entwickelt und implementiert, jegliche Standards des wissenschaftlichen Arbeitens wurden dabei ignoriert. Dies hat eine Gruppe von WissenschaftlerInnen der TU Wien, der WU Wien und der Uni Wien vor Kurzem erst auch recht deutlich zum Ausdruck gebracht:

„Immer wieder wurde seitens des AMS von Transparenz in Bezug auf den AMS-Algorithmus gesprochen. Ob und inwiefern Transparenz in einem produktiven Ausmaß vorliegt, misst sich daran, inwieweit ein wissenschaftlicher Diskurs möglich ist, der auf belegbaren Fakten und Daten basiert. Echte Transparenz würde bedeuten, dass ebendiese belegbaren Fakten und Daten seitens des AMS bereitgestellt werden. Leider wurde dem bisher nicht in zufriedenstellender Weise nachgekommen: Von den 96 Modellvarianten, die sich zu dem algorithmischen System bündeln, wurden nur zwei veröffentlicht, und eine davon erst auf Anfrage. Auch die entsprechenden Fehlerquoten der 96 Modellvarianten sind weitestgehend unbekannt.

Dass es überhaupt eines regen Schriftverkehrs und mehrerer Anfragen bedarf, um auch nur einen Bruchteil der benötigten Informationen zu erhalten, ist unzumutbar und widerspricht dem Postulat der Transparenz. Vielmehr sollte das AMS als Träger von öffentlicher Verantwortung der versprochenen Transparenz eigeninitiativ nachkommen und Anstrengungen unternehmen, die entsprechenden Modellvarianten, Daten und Fakten belegbar, nachvollziehbar und hinreichend anonymisiert aufzubereiten, um eine Analyse im Zuge eines breiten demokratischen Diskurses zu ermöglichen. Die Beurteilung, inwieweit Transparenz für eine ausreichende wissenschaftliche Debatte vorliegt, obliegt der Wissenschaft und kann nicht durch Behauptungen durch Verantwortliche des AMS auf ihren privaten Internet-Seiten ersetzt werden. Derartige Veröffentlichungen auf privaten Kanälen sind nicht überprüfbar und unterliegen keiner angemessenen Kontrolle durch rechtsstaatliche Institutionen. Stattdessen ist die Kommunikation auf den entsprechenden offiziellen Kanälen des AMS, bei der echte Transparenz gelebt werden sollte, zu führen.“³¹

Den AutorInnen des Beitrags ist völlig Recht zu geben: Abgesehen von einer(!) öffentlich zugänglichen Publikation,³² die den methodischen Vor-

gang nur spärlich dokumentiert, sind keine weiteren Informationen erhältlich: Einmal spricht Herr Kopf von „7500 unterschiedlichen Teilmodellen“,³³ ein anderes Mal von „mehr als 90 verschiedenen Algorithmen“³⁴ – wie viele sind es denn nun? Es wurden mit der erwähnten Ausnahme keine Regressionsgleichungen, keine Teststatistiken, keine Daten veröffentlicht. Das Methodenpapier von Synthesis schafft es, keine einzige wissenschaftliche Quelle, die das methodische Vorgehen in einen wissenschaftlichen Diskurs einbetten würde, zu zitieren. Das ist auch eine Leistung – die wahrscheinlich mit dem Selbstverständnis des „Unternehmens Synthesis“ in Zusammenhang steht: Es handelt sich ja lt. Aussagen des langjährigen wissenschaftlichen Leiters bei Synthesis um ein Unternehmen und keine wissenschaftliche Forschungseinrichtung, weshalb das genaue methodische Vorgehen (Schätzgleichungen, Teststatistiken etc.) gewissermaßen dem Betriebsgeheimnis unterliege und daher nicht veröffentlicht werden könne.³⁵

Das gesamte publizierte Papier von Synthesis ist voll von „Dezisionismus“, es werden ständig Gruppen, Kategorien und Merkmale definiert, wobei man sich fragt: Warum so und nicht anders? Warum werden nur zwei Gruppen (jene mit guten und jene mit schlechten Integrationschancen) geschätzt, und die eigentlich wichtige mittlere Gruppe ergibt sich bloß als Residuum? Warum werden die Gruppengrößen nicht nach Regionen ausgewiesen? Warum werden nur Ergebnisse für jenes Drittel an Geschäftsfällen, die am besten dokumentiert werden können, veröffentlicht und nicht auch für die restlichen zwei Drittel mit „partiell validen“ Schätzergebnissen? Hängt es vielleicht damit zusammen, die Treffergenauigkeit zu „verschönern“?³⁶ Am Ende bleibt sogar unklar, ob es sich tatsächlich um Regressionsergebnisse handelt oder ob nur empirisch für einzelne Zellen (Personengruppen mit bestimmten Merkmalen) Werte für die Vergangenheit ermittelt wurden und die Gleichungen erst *ex-post* geschätzt wurden. Darauf gibt es leider keine Antworten, weil diese auch auf Nachfragen nicht zur Verfügung gestellt wurden.³⁷ Anscheinend ist ein Teil dieser Kritik bei Synthesis sogar bereits angekommen, es wird nämlich in einem aktuellen Papier davon gesprochen, betroffene Personen niederschwellig in den Entscheidungsprozess einzubinden, die *Profiling*-Ergebnisse nur als „zweite Meinung“ zu werten und Transparenz herzustellen („Darüber lohnt es, als vertrauensbildende Maßnahme gegenüber der Zivilgesellschaft, Transparenz zu schaffen; über den ‚Algorithmus‘ selbst und die Befunde der Technikfolgenabschätzung.“³⁸) Diese Einsicht kommt spät, sie kommt einseitig, und sie wird von oben herab verkündet – „*ex cathedra*“ gewissermaßen!

Treffergenauigkeit: Die Brauchbarkeit von und das Vertrauen in *Profiling* hängen entscheidend davon ab, wie genau geschätzt werden kann bzw. wie groß der Schätzfehler ist („*false positive*“ und „*false negative*“).³⁹ Syn-

thesis gab bekannt, dass für ca. 440.000 Geschäftsfälle (von insgesamt 1,2 Mio.) voll valide Schätzungen möglich waren, die Trefferquote lag zwischen 80 und 85%.⁴⁰ Allerdings wurden diese Quoten nur für Fälle mit niedriger und mit hoher Integrationswahrscheinlichkeit bekannt gegeben. Es darf vermutet werden, dass diese Trefferquoten niedriger für die mittlere Gruppe sind, aber auch für den großen Rest an Fällen mit lückenhaften Informationen (siehe Anmerkung 36). Wenn also von einer Fehlerquote von 20-40% ausgegangen wird, dann werden 240.000-480.000 Geschäftsfälle oder 180.000-360.000 Personen (von insgesamt ca. 900.000 Personen, die jährlich arbeitslos werden) falsch zugeordnet; d. h. es wird fälschlicherweise angenommen, dass sie in einem bestimmten Zeitraum in einem bestimmten Zeitausmaß beschäftigt sein werden (obwohl sie es nicht sind) und umgekehrt. Führt man sich diese absoluten Zahlen vor Augen, dann müssen starke Zweifel am Nutzen von *Profiling* in der Arbeitsmarktpolitik aufkommen.

Diskriminierung: Diese kann zumindest auf drei Ebenen ansetzen:

(1) Sollte das AMS Personen, die sehr niedrige Chancen zur Wiedereingliederung in den Arbeitsmarkt haben, bestimmte Maßnahmen und Programme vorenthalten, die aber die Beschäftigungsaufnahmen dieser Personen zu vertretbaren Kosten deutlich verbessern würden, dann wäre das ein klarer Fall von Diskriminierung. Bisher hat das AMS aber immer behauptet, dass diese speziellen Problemgruppen am Arbeitsmarkt in Zukunft keine Kurse und Maßnahmen erhalten sollen, die für sie nicht wirken. Genannt wurden in diesem Zusammenhang Facharbeiterintensivausbildungen und Sozialökonomische Betriebe. Das wäre dann aber kein Problem, wobei in dieser sehr wichtigen Debatte die zentrale Kategorie „Effektivität einer Maßnahme für eine bestimmte Gruppe von Arbeitslosen (Langzeitarbeitslose, Jugendliche, Älter, gesundheitlich Beeinträchtigte)“ ist – siehe dazu weiterführend das nachfolgende Kapitel.

(2) Statistisches *Profiling* ist notwendigerweise mit dem Problem der „statistischen Diskriminierung“ verbunden, weil Personen, die einer Gruppe mit bestimmten Merkmalen (Frauen, Niedrigqualifizierte, MigrantInnen etc.) angehören, als ident betrachtet werden, was sie aber naturgemäß nicht sind. Sie weichen vom Durchschnitt etwa in Bezug auf Motivation, Arbeitsmarktnähe, Bildungsaffinität etc. ab.

„Unfortunately, statistical discrimination is an inherent feature of statistical profiling models. Within the framework of a PES, profiling models estimate the probability of work resumption for an individual by relying on average probabilities of the group to which the individual belongs. For instance, if migrants are on average less likely to resume work, than each individual migrant will be assigned a lower probability. In other words, average group characteristics are assigned to an individual. That is exactly the definition of statistical discrimination. As a result, migrants who find a job ex-post are

*more likely to have been wrongly classified ex-ante as ‚high-risk‘ individuals compared to non-migrants. The other side of the coin is that non-migrants are more likely to be wrongly classified as ‚low-risk‘ individuals than migrants. This not only holds for migrants, but for all individual jobseekers belonging to a vulnerable group such as older or low-educated jobseekers. Statistical discrimination cannot be avoided completely, but more accurate models will suffer less from it. However, other types of profiling, particularly case-worker-based profiling, but also rule-based profiling, are also prone to (statistical) discrimination.*⁴¹

(3) Eine Frage, die bisher völlig unter den Tisch gefallen ist bei den Diskussionen, ob *Profiling* zur Diskriminierung bestimmter Arbeitsmarktgruppen führt, ist die folgende: Gibt es bei den AMS-BeraterInnen Vorurteile, diskriminieren sie? Wir wissen dazu interessanterweise überhaupt nichts, außer etwas an anekdotischer Evidenz, dass manche BeraterInnen freundlicher, andere wieder weniger freundlich mit Ihren KundInnen umgehen; das Spektrum soll ja vom Sozialarbeiter bis zum „Quasi-Sadisten“ reichen, hört man. Genaueres ist leider, wie erwähnt, nicht bekannt. Ob schikanöses Verhalten öfter bei Menschen mit Migrationshintergrund, bei gesundheitlich Beeinträchtigten, Frauen mit Betreuungspflichten etc. vorkommt, weil die entsprechenden BeraterInnen Vorurteile haben – das wissen wir leider alles nicht. Ungewöhnlich wäre jedoch, wenn im AMS nur Menschen arbeiten würden, die völlig frei von diesen diskriminierenden Verhaltensweisen wären.

Abschließend zum Thema der Diskriminierung sei noch ein Experte für Arbeitsmarktdiskriminierung, nämlich Sendhil Mullainathan, zitiert, der erst kürzlich in der New York Times meinte:

*„Biased Algorithms Are Easier to Fix Than Biased People – Racial discrimination by algorithms or by people is harmful – but that’s where the similarities end ... Changing algorithms is easier than changing people: software on computers can be updated; the ‚wetware‘ in our brains has so far proven much less pliable.*⁴²

Das Argument läuft im Wesentlichen darauf hinaus, dass in einem Programm durch Anwendung desselben in vielen Durchgängen die diskriminierenden Effekte einfach zu isolieren sind; im Gegensatz dazu werden Menschen sehr ungern zugeben, dass ihr Verhalten diskriminierend ist – weshalb sie dieses zu verbergen suchen werden.

Weitere Kritikpunkte am AMS-*Profiling*-Zugang (Anwendung von Künstlicher Intelligenz ohne sozialwissenschaftliche Begleitung, fehlende Einbindung von Betroffenen, mangelhafter Datenschutzes, Verletzung von Grundrechten durch die Verwendung von Algorithmen) werden in Kapitel 4 behandelt.

3. Strategiewechsel in der Arbeitsmarktpolitik: Es geht um *Targeting*⁴³ und nicht um *Profiling*!

3.1 Wofür möchte das AMS *Profiling* verwenden?

Das ist die zentrale Frage, um die diese Arbeit kreist! Die Antwort lautet: Für die Verbesserung von Effektivität und Effizienz des Maßnahmeneinsatzes. Ist das eine sinnvolle Vorgangsweise? Darauf kann geantwortet werden: Nein! Weil *Profiling* dazu nichts Wesentliches beitragen kann. Für die Beantwortung der Frage, ob ein Programm für bestimmte Arbeitslose wirkt oder nicht – dafür braucht man kein *Profiling*. Das AMS hat also einen Umweg gewählt, der nicht hilfreich ist. Besser wäre es gewesen, das Problem direkt anzugehen: Wie können die Maßnahmen für Arbeitslose mit besonders geringen Reintegrationschancen am Arbeitsmarkt effektiver gestaltet werden? Die bisherigen Aussagen des AMS zu der eingangs gestellten Frage, wofür *Profiling* eigentlich genau verwendet wird, lassen jedoch viele Aspekte im Unklaren – wie nachfolgende Darstellung deutlich machen soll.

„Wir wollen das AMS effizienter machen. Wir forschen schon lang zu verschiedenen Dingen. Es geht darum zu schauen, welche Fördermaßnahmen wie wirken. Wir haben in der Vergangenheit immer wieder auch teure Förderinstrumente für Menschen verwendet, die relativ wenig Output hatten. Auf der anderen Seite haben wir sicher auch immer wieder Förderinstrumente für Menschen genutzt, die mit hoher Wahrscheinlichkeit auch allein eine Arbeit gefunden hätten. Wir werden zusehends gemessen an der Frage, wie effizient wir sind, und auch uns ist das ein Anliegen. Mit modernen wissenschaftlichen Methoden können wir das besser erfüllen ... Die Idee ist, bestehende Förderinstrumente bei unterschiedlichen Gruppen unterschiedlich einzusetzen. Facharbeiterintensivausbildungen sind zum Beispiel in der Gruppe der sehr arbeitsmarktfernen Personen im Verhältnis zu den Kosten wenig effektiv ... Wir haben erfahren, dass wir mit deutlich reduzierten Kosten im Rahmen dieses Programmes [eines Beratungs- und Betreuungsformates, Anmerkung K.B-St.] ähnlich viele Menschen in Beschäftigung bringen wie mit langfristigen und oft teuren Ausbildungsprogrammen.“⁴⁴

„Wir wollen daher Mittel von der Förderung befristeter Beschäftigung [hier sind SÖBs gemeint, Anmerkung K.B-St.] und Qualifizierung für arbeitsmarktferne Personen umschichten zur Förderung von Einrichtungen zur intensiven Betreuung dieser Personen.“⁴⁵

Zum Kerngeschäft des AMS gehört es, die richtigen Arbeitslosen in die richtigen Kurse, Programme und Maßnahmen zu bringen. Diese Konzeption von aktiver Arbeitsmarktpolitik gibt es seit dem AMFG 1968. In diesem Lichte müssen die überraschend offenen Aussagen des AMS-Vorstands Kopf als bestürzend bewertet werden: Er stellt doch tatsächlich nach über fünfzig Jahren Instrumenteneinsatz fest, dass das AMS schauen müsse,

„welche Fördermaßnahmen wie wirken“ und dass „bestehende Förderinstrumente bei unterschiedlichen Gruppen unterschiedlich“ eingesetzt würden. Dass teure Maßnahmen „wenig Output“ brächten. Dass es Mitnahmeeffekte gäbe. Dass „moderne wissenschaftliche Methoden“ bessere Antworten auf diese Fragen liefern würden. Und dass mit „deutlich reduzierten Kosten ... ähnlich viele Menschen in Beschäftigung“ gebracht werden können.⁴⁶

Dazu wäre festzuhalten: Standard-Arbeitsmarktevaluierungen, die etwa *Matching*-Methoden anwenden, gibt es in den USA seit den 1970er-Jahren, aber vielleicht erst kürzlich im AMS. Wenn ein sehr viel billigeres Programm dieselben Wirkungen entfaltet wie ein teures Programm, warum ist man im AMS darauf nicht schon früher gekommen? Gibt es wirklich harte Evaluierungsergebnisse dazu, dass BBEN dieselben Effekte generieren wie SÖBs (es gibt sie nicht, siehe unten), oder sind es nur die üblichen „Brutto-Beschäftigungsaufnahmen“?⁴⁷ Versucht man die Aussagen ernst zu nehmen, dann kann man sich des Eindrucks nicht erwehren, das AMS habe gerade sein Kerngeschäft entdeckt. Eigentlich eine, wie gesagt, bestürzende (Selbst-)Erkenntnis!

Was wurde darüber hinaus vom AMS zur oben gestellten Frage bekanntgegeben?

- Im Zentrum der Förderung soll in Zukunft die Gruppe mit mittleren Integrationschancen stehen, für jene mit niedrigen Chancen soll es nicht mehr alle (teuren) Förderprogramme geben. Die bisherige Zielgruppenlogik soll allerdings beibehalten werden, was u. a. dazu führt, dass alle Jugendlichen in die mittlere Gruppe eingereiht werden. Auch für Ältere über 50 Jahre und Personen mit gesundheitlichen Beeinträchtigungen soll die Einteilung in drei Chancengruppen nicht gelten.
- Arbeitslose mit niedrigen Integrationschancen sollen (in der Pilotphase) zur persönlichen Stabilisierung 12 Monate v. a. durch „Beratungs- und Betreuungsleistungen neu (BBEN)“ extern betreut werden. Dabei müssen diese Personen dem Arbeitsmarkt zur Verfügung stehen, also sich auf Stellenangebote bewerben. Sie können auch andere AMS-Programme besuchen, wenn das „passender und arbeitsmarktpolitisch zweckmäßiger“ erscheint. Wenn es sich bei diesen Personen nicht um arbeitsmarktpolitische Zielgruppen handelt, dann können diese nicht in Sozialökonomischen Betrieben betreut werden; auch Facharbeiterintensivausbildungen sind für sie nicht möglich.⁴⁸
- Im Endeffekt wird sich für ca. 5-10% der Personen mit niedrigen Integrationschancen das Angebot an Förderungen um zwei Maßnahmen (SÖB, Facharbeiterintensivausbildungen) reduzieren.
- Die Letztentscheidung, was mit den Arbeitslosen in den einzelnen Segmenten gemacht wird, bleibt beim Berater bzw. bei der Beraterin; AMAS, also *Profiling*, soll nur eine „zweite Meinung“ darstellen.

- Die Höherstufung von Personen mit niedrigen Integrationschancen durch die BeraterInnen ist immer dann möglich, wenn „durch die in Aussicht genommene Förderung eine entsprechende Erhöhung der Integrationschancen erwartet“ wird.⁴⁹ Eine Herabsetzung der Arbeitsmarktchancen von mittel/hoch auf niedrig kann nur erfolgen, wenn davor ein externer Perspektivencheck durchgeführt wurde (siehe unten).⁵⁰
- Darüber hinaus kann die Förderintensität bei einzelnen Maßnahmen (etwa der Eingliederungsbeihilfe) nach mittelfristigen Integrationschancen abgestuft werden – dabei gilt die Regel: Je geringer die Chance, desto höher die Beihilfe.⁵¹

3.2 Effektivität und Effizienz in der österreichischen Arbeitsmarktpolitik: Bisherige Evaluierungsergebnisse für besondere Problemgruppen

3.2.1 Die grundlegende Problemstellung

Die eigentliche Problemstellung in der Arbeitsmarktpolitik sollte „*Targeting*“ heißen und nicht „*Profiling*“! Es sollte darum gehen, effektive Kurs- und Maßnahmenprogramme zu schaffen, in denen die richtigen Arbeitslosen sitzen, also Personen, bei denen die Nettokosten (Erträge minus Kosten) der Programmteilnahmen maximal sind. Bei der gesamten Diskussion zur Wirksamkeit der aktiven Programme steht im Hintergrund die eigentlich unklar formulierte „Zielfunktion“ (Was soll erreicht werden?) für die AMS-BeraterInnen.⁵² Denn dadurch ergibt sich das Problem, dass nicht klar ist, was der folgende Satz eigentlich bedeuten soll: Für eine arbeitslose Person „zahlt sich eine bestimmte Intervention nicht aus“ – was ist damit genau gemeint? Ist mit zu „teuer“ eigentlich „ineffizient“ gemeint? Gegenwärtig weisen jedenfalls die BeraterInnen ihre KundInnen nach ihren bisherigen Erfahrungen bestimmten Maßnahmen zu, ohne recht zu wissen, welche dieser Interventionen erfolgreich bzw. nicht erfolgreich waren. BeraterInnen haben in der Regel keine Zeit, Evaluierungsstudien zu lesen. Es gibt auch keine Nachverfolgung der Wirkungen von einzelnen Programmen und Kursen. Die BeraterInnen-Entscheidungen sind zudem abhängig von budgetären Gegebenheiten, die sich in der Verfügbarkeit von Kursen ausdrücken, von Zielvorgaben und beispielsweise dem Gefühl, eine Maßnahme wäre etwa zu teuer für einen „Problemfall“, „weil sie ohnehin nichts bringt“. Die Grenzen von Effektivität und Effizienz als die wichtigsten Entscheidungskriterien verschwimmen dabei bis zur Unkenntlichkeit. Was wäre in dieser Situation eine sinnvolle Vorgangsweise? Sollten bestimmte Personen überhaupt keine Maßnahmen mehr bekommen? Sollten die KundInnen segmentiert werden? Diese Fragen lassen sich im

gegenwärtigen Regime nicht eindeutig beantworten. Eine einfache Zuweisungsregel wäre die folgende: „Weise so den Maßnahmen zu, dass die Fiskalbilanz für die Teilnahme an einzelnen Kursen und Programmen aus Sicht der öffentlichen Hand maximiert wird!“ D. h., die wahrscheinlichen Zahlungen für aktive und passive Leistungen (die bei langzeitarbeitslosen Personen sehr hoch sein können) sollen minimiert und die Rückflüsse aus Beschäftigungsaufnahmen sollen maximiert werden. In dieses Kalkül gehen Maßnahmenkosten, die Effektivität einzelner Maßnahmen für einzelne arbeitslose Personen (Beschäftigungstage), wahrscheinliche Kosten für passive Maßnahmen (wie Arbeitslosengeld, Notstandshilfe, etc.) und Rückflüsse für die öffentliche Hand (Steuern/Sozialversicherungsbeiträge für erzielte Einkommen) ein. Bei dieser Regel stellt sich also nicht die Frage, ob für bestimmte Gruppen viel oder wenig unternommen wird; die Zuweisungsregel würde diese Entscheidung übernehmen! Gegenwärtig können BeraterInnen nicht nach dieser Regel entscheiden, weil ihnen schlicht und einfach viele Informationen fehlen.

3.2.2 Beratungs- und Betreuungseinrichtungen (BBEN)

Vor diesem Hintergrund ist die Aussage des AMS-Vorstandes Kopf interessant, dass nämlich mit sehr günstigen Interventionen – gemeint waren neue Beratungs- und Betreuungseinrichtungen, BBEN –, die etwa $\frac{1}{10}$ von teuren Programmen wie SÖB und GBP kosten, ähnlich hohe Wirkungen erzielt werden können. In diesem Zusammenhang stellt sich die Frage, ob es dazu harte Evaluierungsergebnisse gibt? Die Antwort lautet: Nein! Denn diese Aussage bezieht sich auf Beschäftigungsaufnahmen (Stichwort: Bruttoeffekte), die aber nicht mit richtigen Wirkungs-Evaluierungsergebnissen – die nur über Kontrollgruppenanalysen möglich sind; Stichwort: Nettoeffekte – verwechselt werden dürfen. Der Grund ist einfach: Beschäftigungsaufnahmen etwa drei, sechs oder neun Monate nach Maßnahmenteilnahmen können auf Programmwirkungen zurückzuführen sein, sie müssen aber nicht! Sie könnten auch das Resultat von konjunkturellen oder regionalen (eine große Firma geht beispielsweise in Konkurs) oder anderen Einflüssen sein. Erst wenn die Nachkarrieren der TeilnehmerInnen mit jenen der Kontrollgruppen, also den „statistischen Zwillingen“, verglichen werden, kann man von kausalen Wirkungen sprechen.

Diese Fehlinterpretation von Beschäftigungsaufnahmen als Maßnahmenerfolg zieht sich leider durch die alltägliche Praxis des AMS. Jedenfalls bezieht sich die oben zitierte Aussage von Herrn Kopf auf AMS-Analysen⁵³ in einem frühen Stadium der Pilotphase: Diese Bruttoergebnisse zeigten Beschäftigungsaufnahmen drei Monate nach Beendigung der BBEN-Teilnahmen von ca. 33%, was für diese Problemgruppe einen sehr hohen Wert darstellt. Allerdings waren bei mehreren Tausenden Teilnah-

men nur ein Bruchteil (von etwa 20%) „auswertbar“, was Zweifel an der Verlässlichkeit dieser Zahl aufkommen lässt. Spätere Ergebnisse aus dem DWH für die Jahre 2018 und 2019 (3, 6, 9 Monate nach Maßnahmenende) zeigen, je nach gewähltem Zeitraum, weit niedrigere Werte für „nicht geförderte Beschäftigung“ im Bereich von 5 bis 18% aller AbgängerInnen aus BBEN. Werden etwas spezifischer Problemgruppen mit Mehrfachbelastungen (Alter über 45 Jahre, Behinderung, max. Pflichtschule, AL vorangegangenen Arbeitslosigkeitsdauer über ein Jahr) verglichen, die 2018 entweder in einem SÖB oder in BBEN waren, dann zeigt sich für den Bestandwert 6 Monate nach Maßnahmenteilnahme: Nach BBEN-Teilnahme liegt der Wert von allen TeilnehmerInnen bei 7,13% in nicht geförderter Beschäftigung, bei SÖB-Teilnahmen hingegen bei 14,6%.⁵⁴ Dieses Ergebnis widerspricht der Aussage vom AMS-Vorstand Kopf hinsichtlich der Wirkungen, allerdings bleibt die Tatsache, dass SÖBs sehr viel teurer als BBEN sind.

Über diese AMS-Analysen mittels DWH hinaus sind bisher zu den BBEN nur zwei externe Untersuchungen gemacht worden:

- Eine Art *Monitoring* der Maßnahmenumsetzung, die allerdings keine klassische Wirkungsanalyse darstellt.⁵⁵ Es wurden dabei 14 telefonische Interviews mit AMS-LeiterInnen bzw. -BeraterInnen und TrägervertreterInnen, eine *Online*-Befragung mit 375 AMS-BeraterInnen und telefonische Befragungen mit 304 BBEN-TeilnehmerInnen durchgeführt. Dabei ging es für die TeilnehmerInnen in den BBEN, die multiple Vermittlungshindernisse aufwiesen, um die persönliche Stabilisierung, die Stärkung und Aktivierung des Selbsthilfepotenzials, um die Erhöhung der Chancen für die perspektivische (!) Integration in den Arbeitsmarkt und die Unterstützung beim Übergang in das adäquate Sozial- und Versorgungssystem.⁵⁶ Die Teilnahme an BBEN war in der Pilotphase freiwillig, was wahrscheinlich dazu führte, dass besonders motivierte Langzeitarbeitslose teilnahmen. Es kam zu einer Entlastung von AMS-BeraterInnen in Bezug auf besonders schwierige Problemgruppen, die für zwölf Monate in den BBEN verbleiben konnten.
- Und eine Fiskalanalyse des WIFO⁵⁷ zeigt, dass die BBEN-TeilnehmerInnen in der Pilot-RGS länger arbeitslos, weniger in unselbstständiger Beschäftigung und weniger erwerbsaktiv waren als Nicht-TeilnehmerInnen in der Pilot-RGS. Das AMS ersparte sich in der Pilotphase Ausgaben für SÖB, GBP, SÖBÜ und sonstige BBE, diese Kostenersparnis wurde aber durch Mehrausgaben für BBEN deutlich überschritten, sodass es zu einem Mehraufwand von € 168.444 pro BBEN-Zielgruppenperson kam. Dieser negative Wirkungsvergleich für die BBEN könnte mit Anlaufschwierigkeit, aber auch mit dem Umstand zu tun haben, dass die AMS-BeraterInnen für Personen mit multiplen

Vermittlungshemmnissen, die nicht in BBEN waren, mehr Zeit zur Verfügung hatten.

3.2.3 Frühere Evaluierungsergebnisse zur Wirksamkeit von Programmen für besondere Problemgruppen am Arbeitsmarkt

State of the Art-Maßnahmenevaluierungen zeigen ein gemischtes Bild der kausalen Wirkungen von aktiver Arbeitsmarktpolitik in Österreich.⁵⁸ Manche Programme wirken nicht für den Durchschnitt, d. h. etwa im Vergleich einer Nachbeobachtungsperiode von drei Jahren hat die Kontrollgruppe beispielsweise mehr Tage in ungeförderter Beschäftigung und weniger Tage in Arbeitslosigkeit verbracht; oder sie wirken im Durchschnitt der TeilnehmerInnen, mit entsprechenden Effekten auf Beschäftigung, Arbeitslosigkeit und Einkommen. Fast nie wirken die Programme für niemanden, Effektheterogenität ist also häufig der Fall, d. h. die Programme wirken unterschiedlich für Männer und Frauen, Jugendliche und Ältere, gering und höher Qualifizierte, InländerInnen und AusländerInnen, etc.

Frühe Evaluierungen des gesamten Förderprogrammes des AMS⁵⁹ zeigten etwa, dass sich die durchschnittlichen positiven Beschäftigungswirkungen von Qualifizierungsprogrammen erst am Ende einer dreijährigen Nachbeobachtungsperiode einstellen, bei Kurskostenbeihilfen ist dies bereits nach 1,5 Jahren der Fall. Dies gilt auch für Sozialökonomische Betriebe, also öffentliche Beschäftigungsprogramme für Problemgruppen, die, im internationalen Vergleich betrachtet, erfreulich erfolgreich zu sein scheinen. Bei der Eingliederungsbeihilfe wiederum sind etwa 50% Mitnahmeeffekte zu berücksichtigen; diese Förderung wirkt v. a. bei älteren Arbeitslosen und bei Frauen.⁶⁰

Welche Ergebnisse gibt es aber für die in unserem Zusammenhang besonders relevante Gruppe der Personen mit multiplen Problemlagen (Längzeitbeschäftigungslose, Ältere, Behinderte, Unqualifizierte etc.)? Interessanterweise ist die Evaluierungsevidenz zu diesen Gruppen vergleichsweise „dünn“, weil in *Matching*-Analysen die Wirkungen meist nur nach Geschlecht und Alter differenziert werden. Zwei Ausnahmen seien hier kurz zitiert:

- Während eine frühere Untersuchung⁶¹ der Sozialökonomischen Betriebe (SÖB) noch zu dem Ergebnis kam, dass dieses Programm für niedrig Qualifizierte (maximal Pflichtschulabschluss) im Gegensatz zu höher qualifizierten Arbeitslosen durchaus zu positiven Beschäftigungseffekten gegenüber einer Kontrollgruppe führt, kommt eine rezentere Studie⁶² zum gegenteiligen Ergebnis: Hier zeigen sich die stärkeren positiven Effekte bei hohem Bildungsniveau, aber interessanterweise auch bei Langzeitarbeitslosen. Daraus folgt für die Wirkungen von SÖBs: Nach dieser neueren Untersuchung sind diese

Programme nicht sehr wirksam, was die Tage in ungeförderter Beschäftigung betrifft, womit auch ihre Fiskalbilanz negativ ist.

- Für die Gruppe der gesundheitlich beeinträchtigten Arbeitslosen in Oberösterreich haben WIFO/Prospect 2017 eine Untersuchung⁶³ zu den Nettoeffekten der Programmteilnahmen gemacht, mit sehr ähnlichen Ergebnissen wie für alle Arbeitslosen: Die Wirkungen sind insgesamt sehr schwach, sie nehmen aber über den Nachbeobachtungszeitraum hin zu. Wenn das erste Jahr des Förderbeginns (und damit der *Lock-in*-Effekt) vernachlässigt wird – damit werden die Ergebnisse etwas positiver dargestellt –, dann zeigen im 2. und 3. Jahr Kurskosten, Beratungs- und Betreuungseinrichtungen und SÖB/GBP schwach positive Wirkungen, bei den Eingliederungsbeihilfen fallen die Effekte mit +72,8% (oder +160 Tagen) deutlich aus.

3.3 Die Erfahrungen in der Schweiz mit „Statistical Assistance for Programme Selection (SAPS)“

In der Schweiz wurde am Beginn der 2000er-Jahre der Versuch unternommen, die Zuweisung von Arbeitslosen zu Programmen der aktiven Arbeitsmarktpolitik mit Hilfe von statistischen Methoden so zu unterstützen, dass jene Personen in den Maßnahmen sitzen, für welche die Ergebnisse maximal sind.

„One possible technique is based on the idea that we want to compare the employment of one person in programme A with the one of a similar person in programme B, where similar means that the two persons should be identical with respect to all characteristics that matter for their employability as well as their selection into programmes. Conditional on all these characteristics, there is no selection bias. Therefore, conditional on these characteristics, the labour market outcomes of participants in programme A and programme B can be compared to judge the impact of programme A versus B. Such an estimation technique, however, is only applicable if a very rich data set including all variables that affected both programme assignment and labour market outcomes is available.“⁶⁴

In der oben zitierten Studie wird die aktuelle Verteilung auf die Programme verglichen mit einer zufälligen und verschiedenen Varianten von optimaler Verteilung – in Bezug auf die Ergebnisvariablen Beschäftigung und Einkommen. Die zentrale Erkenntnis aus dieser Simulation ist, dass die Beschäftigungsanteile zwölf Monate nach Programmteilnahme mit Hilfe von SAPS von 49,8% auf 57,7% (Ann.: Budgets für einzelne Programme bleiben konstant) bzw. 64,5% gesteigert werden können. Das sind Steigerungen von ca. 8 (bzw. 16%) bzw. ca. 15 Prozentpunkten (30%) – d. h. die Arbeit von BeraterInnen in Bezug auf die Zuweisung zu Programmen kann mit Hilfe von statistischen Methoden deutlich gesteigert werden!

Leider hat sich in einem daran anschließenden Experiment in fünf Regionen gezeigt, dass die BeraterInnen das Hilfsinstrument SAPS nicht wirklich genützt haben: Nur etwa ein Drittel verwendete die Information überhaupt, und nur ca. 12% bzw. 29% (je nach Definition von „*compliance rate*“) hielten sich an die Empfehlung von SAPS, für eine bestimmte Person ein bestimmtes Programm auszuwählen. Damit konnte SAPS natürlich keine Wirkung auf die Zuweisungspraxis haben. Die AutorInnen der Untersuchungen meinen, dass die Angst der BeraterInnen davor, dass der „Computer“ sie in Zukunft ersetzen könnte bzw. er ihre Autonomie einschränken könnte, entscheidend dafür war, warum dieses Instrument in der Praxis nur wenig genützt wurde.⁶⁵ Dies ist ein deutliches Indiz dafür, wie wichtig es ist, die institutionellen Logiken und möglichen Widerstände frühzeitig zu berücksichtigen.

4. Algorithmen, algorithmische Entscheidungssysteme und künstliche Intelligenz in der Arbeitsmarktpolitik

„An algorithm is a set of instructions – a preset, rigid, coded recipe that gets executed when it encounters a trigger. AI on the other hand – which is an extremely broad term covering a myriad of AI specializations and subsets – is a group of algorithms that can modify its algorithms and create new algorithms in response to learned inputs and data as opposed to relying solely on the inputs it was designed to recognize as triggers. This ability to change, adapt and grow based on new data, is described as ‚intelligence‘.“⁶⁶

Algorithmen können also als Handlungsanweisungen beschrieben werden, die von der Eingabe (also den Ausgangsinformationen) zu einem Ergebnis bzw. zur Ausgabe mit gewünschten Eigenschaften führen.⁶⁷ Ein Algorithmus kann etwa anhand von Straßenkarten, Start- und Zielpunkt und der aktuellen Verkehrslage einen Weg zum Ziel mit der frühesten Ankunftszeit finden. Aber Algorithmen werden auch eingesetzt, um beispielsweise herauszufinden, wie hoch die Wahrscheinlichkeit ist, dass ein Gefängnisinsasse rückfällig wird. Oder wie wahrscheinlich jemand einen Unfall verursacht.

„Das Problem ist, dass der Algorithmus die Daten von Herrn Schmidt und Herrn Müller nimmt, um vorherzusagen, ob Herr Meier rückfällig werden wird. Bei Amazon beispielsweise ist das anders. Hier berechnet der Computer die Empfehlungen basierend auf Ihrer eigenen Kaufhistorie. Das halte ich für fair. Bei der Rückfallprognose jedoch bestimmt man das persönliche Risiko eines Straftäters aus dem Verhalten anderer Menschen, die ihm in einigen Aspekten ähneln. Ich nenne das ‚algorithmische Sippenhaft‘ ...“; „... (Frage: Ein Psychologe würde den Job also besser machen?) Das kann man so pauschal nicht sagen. Der Vorteil des Algorithmus ist es nämlich, dass er vollkommen objektiv, sprich bei derselben Sachlage immer gleich entschei-

det. Das Urteil eines menschlichen Gutachters ist dagegen leicht zu beeinflussen. Es hängt unter anderem davon ab, ob er gerade gegessen, ob er sich über irgendetwas geärgert oder ob seine Ehefrau ihm vor der Arbeit noch einen Abschiedskuss gegeben hat. Das spielt beim Computer keine Rolle.“⁶⁸

Algorithmische Entscheidungssysteme (ADM) wiederum beinhalten und erzeugen Regeln, nach denen Entscheidungen getroffen werden können.

„Algorithmische Entscheidungssysteme beinhalten Algorithmen an zwei Stellen: Der erste Algorithmus lernt auf Basis der Daten ein statistisches Modell. Das statistische Modell ist dann die Grundlage für den (meist sehr einfachen) zweiten Algorithmus, der die eigentliche Entscheidung für eine neue Eingabe berechnet ... Das Ergebnis eines algorithmischen Entscheidungssystems ist das Produkt aus der Interaktion von Daten und dem ersten Algorithmus. Man muss also das Gesamtsystem betrachten, bestehend aus Daten, dem ersten Algorithmus, der das Modell lernt, und dem Modell, das dann die Grundlage für die Entscheidungen bietet.“⁶⁹

Als Beispiel für ein algorithmisches Entscheidungssystem führt Katharina Zweig Daten von AutofahrerInnen an (etwa Alter, Geschlecht, Anzahl der selbstverschuldeten Unfälle), für die mithilfe des oben beschriebenen statistischen Modells eine Regel aufgestellt wird, die die Menge aller AutofahrerInnen in zwei eindeutig separierte Gruppen teilt: „Verursacht wahrscheinlich einen Unfall“, „Verursacht wahrscheinlich keinen Unfall“. Weitere Beispiele für ADM-Systeme sind: Googles Suchmaschinenalgorithmus, Facebooks Newsfeed, „*predictive policing*“ (d. h., dass etwa aus bisherigen Straftaten abgeleitet wird, wann wo welche Straftaten zu erwarten sind), Rückfälligkeitsvorhersagealgorithmen, Terroristenidentifikation.⁷⁰ Zweig nennt darüber hinaus mehrere Fehlerquellen bzw. Fehlrurteile von ADM-Systemen: Zufällige Faktoren, wie Erkrankungen, Teamatmosphäre; zu kleine und zu inhomogene Datenmengen, etwa bei der Identifikation von Terroristen; fehlerhafte Entwicklung und fehlerhafter Einsatz von ADM-Systemen. Zudem – und das ist ein wesentlicher Aspekt – können Daten schon zu Beginn Diskriminierungen enthalten, beispielsweise wenn Frauen oder MigrantInnen zu selten zu Bewerbungsgesprächen eingeladen werden; in diesem Fall wird die Diskriminierung vom Algorithmus „mitgelernt“. Am Beispiel der sozialen Medien wird darüber hinaus deutlich, dass die Operationalisierung des schwer für den Computer fassbaren Begriffes „Relevanz“ eine große Herausforderung darstellt. Bei selbstlernenden ADM-Systemen kommt noch ein anderer wichtiger Aspekt hinzu: Es ist nämlich in diesen Fällen das Zustandekommen der Ergebnisse in der Regel nicht nachvollziehbar.

„ADM-Systeme sind folglich fehleranfällig, und oft ist es schwierig, bei Fehlentscheidungen die dafür Verantwortlichen zu identifizieren. Die Forderung nach einer Qualitätssicherung und mehr Transparenz ist daher berechtigt.“⁷¹

Zweig schlägt demzufolge fünf Risikoklassen von ADM-Systemen vor, bei denen von Klasse 0 bis Klasse 4 die Überwachungs-, Kontroll- und Transparenzpflichten stetig verschärft werden. Bei Klasse 1 wären beispielsweise für die Google-Suchmaschine *Black-Box-Analysen* ohne Zugriff auf den Code vorzusehen. Als Beispiele für Klasse 2 sind automatische oder unterstützende ADM-Systeme, die Bewerbungen bewerten, zu nennen. Für diese sollten ständige Kontrollen und mehrere Transparenzpflichten festgelegt werden. In Klasse 3 dürfen nur gut nachvollziehbare, erklärende Modelle verwendet werden (etwa bei der Klassifizierung des zukünftigen Erfolges von ArbeitnehmerInnen). Als Beispiele für Klasse 4 wären die automatische Tötung von gesuchten Personen oder die Identifikation von Terroristen zu nennen – in diesen Fällen sollten ADM-Systeme mit lernenden Komponenten gänzlich verboten werden.

Um die Fehlerquellen und die Risiken bei der Entwicklung und der Einbettung von algorithmischen Entscheidungssystemen zu minimieren, schlägt Katharina Zweig⁷² einen Prozess in fünf Phasen vor:

1. Phase: Algorithmen werden entworfen und in Software implementiert;
2. Phase: Optional erfolgt die Operationalisierung, es werden also Konstrukte wie „Relevanz einer Nachricht“ in Indikatoren übersetzt (z. B. Häufigkeit des Anklickens); und Auswahl der Daten, mit denen der Algorithmus trainiert werden soll und Auswahl der Bewertungs- oder Vorhersagemethode;
3. Phase: Hier werden eine Methode des maschinellen Lernens mit den Trainingsdaten zusammengebracht und das Entscheidungssystem konstruiert;
4. Phase: Einbettung in die gesellschaftliche Praxis, das System wird auf neue Daten angewendet, die Ergebnisse anschließend interpretiert und in einer Handlung umgesetzt;
5. Phase: Evaluation des Entscheidungssystems.

Was bedeuten diese Ausführungen zu Algorithmen und algorithmischen Entscheidungssystemen (als eine sehr schwache Form von künstlicher Intelligenz) in Bezug auf das vom AMS gewählte *Profiling-System*?

- Beim „AMS-Chancen-Modell“ werden mittels logistischer Regressionen die Wiederbeschäftigungswahrscheinlichkeiten von arbeitslosen Personen ermittelt und die Arbeitslosen anschließend in drei Gruppen eingeteilt (hoch, mittel, niedrig). Diese Gruppeneinteilung wurde von außen mehr oder minder willkürlich (es hätten auch 5 oder 2 oder 10 Gruppen sein können) vorgegeben.
- Daraus folgt: Das AMS-Chancen-Modell ist kein Algorithmus – in diesem Fall würde die „Maschine“ eindeutig festlegen, welche Risikogruppen in welche Maßnahmen kommen (das passiert aber nicht, Stichwort: Zielgruppenlogik). Es ist aber auch kein algorithmisches Entscheidungssystem, denn in diesem Fall käme die Gruppeneinteilung

lung oder ein anderer Vorschlag zur Aufteilung der Arbeitslosen ebenfalls von der „Maschine“.

- Und natürlich handelt es sich beim AMS-Chancen-Modell nicht um künstliche Intelligenz (was eigentlich schade ist; siehe unten).
- Aber dennoch sollten die Vorschläge von Katharina Zweig u. a. zur Entwicklung von ADM-Systemen und zu ihrer Einteilung in Risikoklassen mit den entsprechenden Überwachungs-, Kontroll- und Transparenzpflichten auch beim AMS-Chancen-Modell ernst genommen werden bzw. geprüft werden, welche Relevanz ihnen dabei zukommt.

4.1 Künstliche Intelligenz in der Arbeitsmarktpolitik

Wie oben bereits in Bezug auf selbstlernende algorithmischen Entscheidungssystemen (ADM) ausgeführt wurde, erleichtert auch künstliche Intelligenz die Datenverarbeitung und insbesondere die Anpassung an neue Umstände. Aber sie benötigt auch extrem viele Daten, und sie verwendet eine Menge an *Black Box*-Techniken, und: *„These methods have an innate tendency to perpetuate and amplify biases present in the data with respect to factors such as race, gender, and education.“*⁷³

Unabhängig von diesen kritischen Aspekten von künstlicher Intelligenz gilt aber auf einer grundsätzlicheren Ebene: Die Arbeitsmarktpolitik mit ihrem Übermaß an verschiedensten Administrativ-Daten zu den Arbeitslosen – Stichwort: *Big Data* – wäre eigentlich ein ideales Anwendungsgebiet für künstliche Intelligenz, im Sinne von „Maschinenlernen“. Von BeraterInnen kann die Vielzahl von Daten und Informationen keinesfalls strukturiert verarbeitet werden, weil sie immer auch nur einen kleinen Ausschnitt der Wirklichkeit zu sehen bekommen – sie kennen die Vorgeschichte der Person, die vor ihnen sitzt, nur rudimentär, sie kennen die Wirkungen ihrer Programme nur ansatzweise, und sie haben nur ein eingeschränktes Wissen über die Konsequenzen ihrer Handlungen. Daher sollte „die Maschine“, also ein EDV-Programm, mit folgenden Daten gefüttert werden:

- a) Arbeitsmarkthistorie der arbeitslosen Person;
- b) gegenwärtige Situation der betreffenden Person und auf dem regionalen Arbeitsmarkt;
- c) Evaluierungsergebnisse für die Maßnahmenteilnahme von vergleichbaren Personen (Fiskalbilanz, Effektivität)
- d) Beraterentscheidungen in vergleichbaren Fällen.

Künstliche Intelligenz würde diese Informationen verwenden und der/dem Berater/in einen Vorschlag für eine Maßnahme welcher Art auch immer machen. Das „Lernen“ der Maschine bestünde nun darin, dass richtige (im Sinne von erfolgreicher Integration) und falsche (nicht erfolgreiche Wiedereingliederung) Informationen wieder eingespeist werden in das Programm und die „Maschine“ dann in der Folge zukünftige ähnliche Ent-

scheidungen entsprechend adaptiert. Das Programm würde also aus den vergangene Entscheidungen lernen. Erste Ansätze zur Anwendung von schwacher künstlicher Intelligenz wurden oben in Bezug auf Flandern und die Niederlande erwähnt.

Soweit die Theorie. In der Praxis ist das österreichische AMS meilenweit von dieser möglichen Anwendung von schwacher künstlicher Intelligenz entfernt. Und angesichts der Interessenlage der beteiligten Akteure wird das wohl auch noch lange so bleiben. Eigentlich schade.

5. Schlussfolgerungen

Aus den vorangegangenen Ausführungen zum AMS-*Profiling* (AMAS) können folgenden Schlüsse gezogen werden:

- Das AMS sollte besser heute als morgen das teure und unnötige Experiment mit dem *Profiling* beenden, weil es zur eigentlichen Frage – Wie können Effektivität und Effizienz des Maßnahmeneinsatzes in der aktiven Arbeitsmarktpolitik verbessert werden? – nichts beitragen kann.
- Die internationalen Erfahrungen mit *Profiling* seit den frühen 1990er-Jahren zeigen, dass eine Vielzahl von Problemen mit diesem Verfahren verbunden ist: Die vorgenommenen Schätzungen zur Wiedereingliederungswahrscheinlichkeit der Arbeitslosen sind zu ungenau, es gibt erhebliche institutionelle Widerstände bei der Anwendung von *Profiling* (etwa in den USA und in der Schweiz), und vor allem: Das *Profiling* hilft wenig bei der Frage, welche Programme für die besonders benachteiligten Arbeitsmarktgruppen wirksam sind.
- Das AMS hat in Österreich ein Institut zur Entwicklung und Implementierung des *Profiling* ausgewählt, das in der Szene der arbeitsmarktpolitischen Forschung bekannt für seine intransparente Vorgangsweise ist. Diesen Mangel haben mehrfach ForscherInnen kritisiert, leider wurden daraus von Seiten des AMS oder der Aufsichtsbehörde, dem BMAFJ, noch keine Konsequenzen gezogen.
- Von Seiten des AMS wurde argumentiert, dass *Profiling* einen Strategiewechsel in der Arbeitsmarktpolitik unterstützen sollte. Bisher ist allerdings von einem derartigen Wechsel nichts zu erkennen, außer dass einige Arbeitsmarktgruppen mit besonderen Vermittlungshemmnungen mehr Beratung und Unterstützung (BBEN) von Trägern außerhalb des AMS erhalten und teurere Kurse nicht mehr bekommen sollen. Dafür hätte es allerdings das *Profiling* nicht gebraucht.
- Richtig ist allerdings, dass das AMS bisher bei der Vermittlung von Problemgruppen des Arbeitsmarktes nicht besonders erfolgreich war – darauf verweisen Ergebnisse von Maßnahmenevaluierungen.

- Die Erfahrungen mit „*Statistical assisted programme selection*“ (SAPS) in der Schweiz böten eine mögliche Orientierung, um die Kurse und Programme besser auf diese schwierigen Gruppen abzustimmen.
- Was das AMS bisher in Bezug auf *Profiling* gemacht hat, hat nichts mit Algorithmen und künstlicher Intelligenz zu tun. Aber die Voraussetzungen für den Einsatz von künstlicher Intelligenz (Stichwort: *Big Data*) in der Arbeitsmarktpolitik wären gut.
- Dies allerdings nur, wenn anders als bisher vorgegangen wird: Dazu wäre es unbedingt notwendig, die Entwicklung und Implementierung von künstlicher Intelligenz mit einer umfassenden sozialwissenschaftlichen Begleitung und Technologiefolgenabschätzung zu versehen, die BeraterInnen und die Arbeitslosen als die Betroffenen von Beginn an einzubinden und ethische und menschenrechtliche Standards zu berücksichtigen. Begleitet und kontrolliert sollte ein derartiger Prozess von einer unabhängigen ExpertInnengruppe, die dem Parlament rechenschaftspflichtig sein sollte.
- Gerade in Zeiten wie der gegenwärtigen Corona-Krise tritt die überragende Bedeutung von aktiver Arbeitsmarktpolitik deutlicher hervor als in ruhigeren Zeiten: Der drastische Anstieg der Arbeitslosigkeit, insbesondere auch der Langzeitarbeitslosigkeit, führt uns vor Augen, wie notwendig die effektive Unterstützung v. a. der besonders vulnerablen Gruppen am Arbeitsmarkt ist. Dazu sind bewährte Instrumente und eine offene Diskussion über weniger hilfreiche Programme und Maßnahmen erforderlich.

Anmerkungen

- ¹ Wenn in diesem Artikel von *Profiling* gesprochen wird, dann ist damit „statistisches *Profiling*“ gemeint. Dabei werden meist mittels logistischer Regressionen Wahrscheinlichkeiten geschätzt, langzeitarbeitslos zu werden, den Arbeitslosenanspruch aufzubrauchen oder in einer bestimmten Zeit wieder einen Arbeitsplatz zu finden. Jedenfalls handelt es sich dabei um ein empirisch-statistisches Datum. Andere Formen von *Profiling* sind: „*Caseworker-based profiling*“, hier kommt dem/der BeraterIn eine zentrale Stellung zu (etwa in Estland und Deutschland); „*Rules-based profiling*“, hier werden einfache Regeln angewendet, etwa, dass Personen mit einem bestimmten Alter oder einer bestimmten Arbeitslosigkeitsdauer in ein Programm kommen können (siehe Norwegen und Polen); schließlich „*data-assisted profiling*“, das ist eine Mischform aus BeraterInnen-Entscheidungen und statistischer Unterstützung; Loxha und Morgandi (2014) 13.
- ² Ein Algorithmus ist eine Abfolge von Befehlen, eine Handlungsanleitung, eine Zuordnungsregel ähnlich einem Kochrezept: „Man nehme ...“. Siehe weiterführend im Detail unter 4.1.
- ³ Beispielsweise Berger et al. (2001).
- ⁴ Szigetvari (2019), Kopf (2019a, c).
- ⁵ Für einen frühen Versuch zur Bestimmung der individuellen Dauer der Arbeitslosigkeit im österreichischen AMS siehe Dornmayr (1996).

- ⁶ OECD (1998). Siehe auch PLS Ramboll (2001): Hier steht vor allem der präventive Aspekt im Vordergrund, *Profiling* soll also dabei helfen, dass Langzeitarbeitslosigkeit erst gar nicht entsteht.
- ⁷ OECD (1998) 9.
- ⁸ Diese Feststellung ist keinesfalls trivial, denn sie bezieht sich auf den besonders reichen Datensatz im Bundesstaat Kentucky; in anderen Bundesstaaten waren die Prognosen weniger erfolgreich.
- ⁹ Diese Wiedereingliederungsmaßnahmen beinhalten v. a. weniger intensive Beratungs-, Betreuungs- und Vermittlungsaktivitäten. Nur 13,8% der BezieherInnen von Arbeitslosengeld wurden zu teureren und intensiveren Programmen wie Qualifizierung und Training vermittelt. Black et al. (2003) 1315, 1317.
- ¹⁰ Für Großbritannien liegen aus dieser Zeit ähnliche Befunde vor; OECD (1998) 21.
- ¹¹ Wie unterschiedliche Funktionsformen bzw. Schätzmethoden (Logit, Probit, Tobit und OLS); unterschiedliche *Outcome*-Variablen und „cut off points“, d. h. angenommene Entscheidungsregel, dass 30%, 50% bzw. 70% der nach geschätzten Wahrscheinlichkeiten geordneten Personen eine Maßnahme erhalten; Modelle mit unterschiedlicher Anzahl an erklärenden Variablen (von „*parsimonious*“ bis „*full modell*“, d. h. etwa 220 Variablen); Modelle für einzelnen Untergruppen von Personen.
- ¹² Die individuelle Wahrscheinlichkeit, keine Beschäftigung 1 Jahr nach der Beobachtungsperiode im Juni/Juli 2000 zu haben; die Wahrscheinlichkeit des Bezuges von Transferleistungen 1 Jahr nach Beobachtungsperiode; die Wahrscheinlichkeit des Prozentanteils an Leistungsbezug in einer Zeitperiode von 30 Monaten nach Beobachtungsperiode.
- ¹³ Für einen aktuellen Überblick zur Debatte siehe etwa Loxha und Morgandi (2014), Kurekova (2014), DG Employment, Social Affairs and Inclusion (2015, 2018) und Desiere et al. (2019).
- ¹⁴ Desiere (2019) 6f.
- ¹⁵ Langenbucher (2019).
- ¹⁶ Desiere et al. (2019) 13.
- ¹⁷ „*Statistical models are never perfectly accurate. Some jobseekers will be wrongly classified in the high-risk group and nevertheless quickly resume work, whereas other will be classified as low-risk and become long-term unemployed. Both errors reduce cost-efficiency of service delivery and, perhaps more importantly, jeopardise confidence and trust in the use of profiling.*“ Desiere et al. (2019) 15.
- ¹⁸ Desiere (2019) 24.
- ¹⁹ Gibbons (2018) 8. Siehe auch Loxha und Morgandi (2014) 20.
- ²⁰ OECD (2018) 3.
- ²¹ Dornmayr (1996), Synthesis (2007).
- ²² Synthesis (2016ab, 2017, 2018, 2019).
- ²³ Synthesis (2018).
- ²⁴ Die nachfolgende Darstellung und die Zitate beziehen sich auf Synthesis (2018). Genau genommen handelt es sich um eine Vielzahl von Modellen; siehe weiterführend unten. Im AMS wird das *Profiling* als „AMAS“ bezeichnet, die Abkürzungen für „Arbeitsmarktchancen-Assistenz-System“.
- ²⁵ Bei Personen mit lückenhaften Informationen (etwa Jugendliche, MigrantInnen, und Arbeitslose mit fragmentierten Erwerbskarrieren) sind nur partiell valide Schätzungen möglich; das sind ca. 760.000 Geschäftsfälle; Synthesis (2018) 4, 15.
- ²⁶ Ein inhärentes Problem dieser Grenzen besteht darin, dass sich Personen mit ähnlichen Integrationswahrscheinlichkeiten, etwa mit 65% und 66% kurzfristiger und 24% und 25% langfristiger Wahrscheinlichkeit, in ihren Arbeitsmarkteigenschaften sehr wenig unterscheiden, aber die Konsequenzen der Gruppeneinteilung doch erheblich sein können.

- ²⁷ Synthesis (2018) 11.
- ²⁸ Junger Mann, maximal Pflichtschule, Österreicher, keine Betreuungspflichten, nicht gesundheitlich beeinträchtigt, wohnhaft etwa in einem Arbeitsmarktbezirk wie Bregenz oder Amstetten, vor Arbeitslosigkeit beschäftigt im Dienstleistungssektor.
- ²⁹ Vergleich Synthesis (2019a) 16 mit Synthesis (2018) 15.
- ³⁰ Kopf (2019c).
- ³¹ Cech et al. (2019); und aktuell auch Allhutter et al. (2020).
- ³² Synthesis (2018).
- ³³ Kopf (2019c).
- ³⁴ Kopf (2018b).
- ³⁵ Diese Diskussionen mit Synthesis wurden schon seit langer Zeit geführt und sind unter ArbeitsmarktforscherInnen hinreichend bekannt. Unverständlich in diesem Zusammenhang ist der Umstand, dass dieses Unternehmen eine europaweite Ausschreibung (die immerhin € 237.142,80 ausmachte) gewinnen konnte, bei der so namhafte Forschungsinstitute wie etwa das WIFO gerade bei der „wissenschaftlichen Eignung“ schlechter als Synthesis abgeschnitten haben; siehe Wimmer (2018). Aufgrund einer parlamentarischen Anfrage mussten die Zahlen für die Entwicklungskosten offensichtlich kräftig nach oben korrigiert werden, nämlich auf € 1,8 Mio.; siehe Szgetvari (2020).
- ³⁶ Tatsächlich liegen die Schätzergebnisse für „partiell valide“ Fällen (Jugendliche, Migrationshintergrund, fragmentierte Beschäftigung) deutlich unter den 80-85%, nämlich teilweise zwischen 50-63%; Synthesis (2017) 33, 35.
- ³⁷ So ist bei Lopez (2019) 12 nachzulesen: „*Other coefficients for other model variants could not be found, and a request to Synthesis for even example coefficients for f_3 was rejected.*“
- ³⁸ Synthesis (2019b) 10.
- ³⁹ In einem Schreiben an die Gleichbehandlungsanwaltschaft hält das AMS fest: „Die statistisch übliche Differenzierung der falsifizierten Prognosen in „*false positives*“ und „*false negatives*“ macht bei der Bestimmung der Treffsicherheit des Arbeitsmarktchancen-Modell des AMS keinen Sinn. Die Grenzwerte für die Einstufung in die 3 Chancensegmente wurden bei beiden Indikatoren nur in eine Richtung wirkend festgelegt (Kurzfristindikator: „größer als“ bzw. Mittelfristindikator: „kleiner als“). Damit kann es bei der Bestimmung der Arbeitslosen mit hohen Chancen nur „*false positives*“-Prognosen geben, also Personen, die trotz positiver Prognose, das Integrationsziel nicht erreichen. Bei der Gruppe mit niedrigen Chancen kann es nur „*false negatives*“-Prognosen geben, also Personen, die trotz negativer Prognose das Integrationsziel doch erreichen. Die Gruppe mit mittleren Integrationschancen ist einfach die Restgröße für die es weder eine „positive“ noch eine „negative“ Integrationsprognose, daher auch keine Falsifikation dieser Prognose gibt.“ Buchinger (2019).
- ⁴⁰ Synthesis (2018) 15. „Im internationalen Vergleich ist das eine sehr hohe Trefferquote (siehe die vorangegangene Tabelle 1), wobei Zweifel berechtigt sind, warum gerade Österreich so viel besser abschneidet als viele andere Länder.“ Buchinger (2019), Frage 13.
- ⁴¹ Desiere (2019) 24.
- ⁴² Sendhil Mullainathan, The New York Times (6.12.2019).
- ⁴³ Eberts et al. (2002).
- ⁴⁴ Kopf (2018).
- ⁴⁵ Kopf (2019a).
- ⁴⁶ Es ist die Rede von $\frac{1}{10}$ der Kosten bei „Beratungs- und Betreuungseinrichtungen neu“ (BBEN) im Vergleich etwa zu SÖBs; siehe BMASGK (2019) 219 – dort werden Ausgaben pro TeilnehmerInnen bei SÖBs von ca. € 7000 im Vergleich zu BBE von € 560 angegeben.

- ⁴⁷ In diesem Falle werden Beschäftigungsaufnahmen nach Maßnahmenende mit kausalen Wirkungen des Programmes verwechselt; siehe weiterführend unter 3.2.1.
- ⁴⁸ AMS-Verwaltungsrat (2019), Szigetvari (2019).
- ⁴⁹ AMS-Verwaltungsrat (2019) 3.
- ⁵⁰ Siehe Liebeswar et al. (2019).
- ⁵¹ Buchinger (2019) Frage 7.
- ⁵² Diese Unklarheit scheint auch Resultat von unklaren rechtlichen Regelungen im § 31 AMSG 1994 zu sein. Hier werden zwei grundlegende Ziel vorgegeben, die nicht immer im gleichen Ausmaß erreicht werden können; nämlich jenes der Chancengleichheit (Abs. 3) und jenes der „Sparsamkeit, Wirtschaftlichkeit und Zweckmäßigkeit“ (Abs. 5). Die nachfolgend formulierte Zuweisungsregel versucht diesem *Trade-off* gerecht zu werden.
- ⁵³ Weinberger (2019).
- ⁵⁴ Das ist zwar keine Kontrollgruppenanalyse, aber es werden zum selben Zeitpunkt dieselben Problemgruppen für zwei Maßnahmen analysiert. Das kommt einem *Matching* recht nahe.
- ⁵⁵ Weber et al. (2019).
- ⁵⁶ TeilnehmerInnen haben multiple Vermittlungshindernisse, d. h. die Geschäftsfall-Nettodauer beträgt mindestens zwei Jahre, in Verbindung mit mindestens zwei der nachfolgenden Kriterien: älter als 44 Jahre, maximal Pflichtschule, gesundheitliche Einschränkungen; Weber et al. (2019) 5.
- ⁵⁷ Böheim et al. (2020). In dieser Untersuchung wurden die Zielgruppen mit multiplen Vermittlungshemmnissen in Pilot-Regionen mit BBEN und ohne BBEN 2018-2019 verglichen in einem Zeitraum von 1,5 Jahren.
- ⁵⁸ Hier sollen nur Evaluierungen behandelt werden, die mit Kontrollgruppen (*Matching*) arbeiten. Für einen Überblick siehe etwa Eco-Austria (2016) und Schweighofer (2013).
- ⁵⁹ Lutz et al. (2005), Lechner et al. (2007).
- ⁶⁰ Eppel et al. (2011).
- ⁶¹ Lechner et al. (2007) 56.
- ⁶² Eppel et al. (2014) 58.
- ⁶³ Eppel et al. (2017).
- ⁶⁴ Behncke et al. (2010) 6.
- ⁶⁵ Behncke et al. (2010) 18.
- ⁶⁶ Ismail (2018). Wobei hier anzumerken wäre, dass wir von „starker“ künstlicher Intelligenz (KI) noch sehr weit entfernt sind. Gegenwärtig finden v. a. Anwendungen mit schwacher künstlicher Intelligenz statt, d. h. diese Art von KI-Systeme funktionieren reaktiv auf einem oberflächlichen Intelligenz-Level und erlangen kein tieferes Verständnis für die Problemlösung. Schwache künstliche Intelligenz ist vor allem auf die Erfüllung klar definierter Aufgaben ausgerichtet und variiert die Herangehensweise an Probleme nicht.
- ⁶⁷ Zweig (2019a) 3.
- ⁶⁸ Zweig (2019b) 60f.
- ⁶⁹ Zweig (2019a) 4.
- ⁷⁰ Lischka und Klingel (2017), hier sind weitere internationale Fallbeispiele für algorithmische Entscheidungssysteme zu finden.
- ⁷¹ Zweig (2019a) 7.
- ⁷² Zweig (2018).
- ⁷³ Niranjan Krishnan, Tiger Analytics, zitiert nach Ismail (2018).

Literatur

- Alhutter, D., Cech, F., Fischer, F., Grill, G., Mager, A., Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective, *frontiers in Big Data*, Volume 3, Article 4 (February 2020). <https://www.frontiersin.org/articles/10.3389/fdata.2020.00005/full>.
- Alston, Ph., Report of the Special rapporteur on extreme poverty and human rights (11 October 2019).
<http://www.statewatch.org/news/2019/nov/un-report-digital-welfare-states-10-19.pdf>.
- AMS-Verwaltungsrat, Ergänzung zum TOP 13, Verwaltungsratssitzung am 10.12.2019, Tischvorlage, Fördermatrix – AMAS-Strategie (Wien, 10. Dezember 2019).
- Arni, P., Schiprowski, A., Die Rolle von Erwartungshaltungen in der Stellensuche und der RAV-Beratung, Teilprojekt 2: Pilotprojekt Jobchancen-Barometer, in: IZA Research Report No. 70 (Dezember (2015)).
<https://www.iza.org/publications/r/178/die-rolle-von-erwartungshaltungen-in-der-stellensuche-und-der-rav-beratung-teilprojekt-2-pilotprojekt-jobchancen-barometer>.
- Behncke, St., Frölich, M., Lechner, M., Statistical Assistance for Programme Selection – For a Better Targeting of Active Labour Market Policies in Switzerland (St. Gallen, September 2010).
https://www.alexandria.unisg.ch/33999/1/Statistical_Assistance_for_Programme_Selection_For_a_Better_Targeting_of_Active_Labour_Market_Policies_in_Switzerland.pdf.
- Berger, M. C., Black, D., Smith, J. A., Evaluating profiling as a means of allocating government services, in: Lechner, M., Pfeffer, F., *Econometric Evaluation of Labor Market Policies*, (Heidelberg, Physica 2001) 59-84.
- Black, D., Smith, J. A., Berger, M. C., Noel, B. J., Is the Threat of Reemployment Services More Effective Than the Service Themselves? Evidence from Random Assignment in the UI System, in: *AER*, September (2003a), 1313-1327.
https://www.ssc.wisc.edu/~scholz/Teaching_742/Black_Smith_Berger_Noel.pdf.
- Black, D., Smith, J. A., Plesca, M., Shannon, S., Profiling UI Claimants to Allocate Reemployment Services: Evidence and Recommendations for States (2003b).
https://wdr.doleta.gov/opr/fulltext/eta_default.cfm?year.
- BMASGK, Aktive Arbeitsmarktpolitik in Österreich 2014-2019, Dokumentation (Wien 2019).
<https://broschuerenservice.sozialministerium.at/Home/Download?publicationId=447>.
- Böheim, R., Eppel, R., Mahringer, H., Kosten-Ertrags-Analyse der „Beratungs- und Betreuungsleistungen für Personen mit multiplen Vermittlungshindernissen“ (BBEN), Präsentation im Strategie- und Förderausschuss des AMS (Wien, 12.2.2020).
- Bryson, A., Kasparova, D., Profiling benefit claimants in Britain: A feasibility study (2003).
<https://webarchive.nationalarchives.gov.uk/20090606030926/http://www.dwp.gov.uk/asd/asd5/rports2003-2004/rrep196.asp>.
- Buchinger, H., Antwortschreiben an die Gleichbehandlungsanwaltschaft (Wien, 24.4.2019).
<https://epicenter.works/sites/default/files/gbaarbeitsmarktchancen-redacted.pdf>.
- Cech, F., Fischer, F., Human, S., Lopez, P., Wagner, B., Dem AMS-Algorithmus fehlt der Beipackzettel, in: *futurezone* (3.10.2019).
<https://futurezone.at/meinung/dem-ams-algorithmus-fehlt-der-beipackzettel/400636022>.
- Council of Europe, Algorithms and Human Rights, Study on the human rights dimension of automated data processing techniques and possible regulatory implications (Strasbourg, March 2018). <https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html>.
- Datenethikkommission, Gutachten der Datenethikkommission der deutschen Bundesregierung, Kurzfassung (Berlin, Oktober 2019).

- https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf?__blob=publicationFile&v=6.
- Desiere, S., Langenbucher, K., Struyven, L., „Statistical profiling in public employment services: An international comparison“, OECD Social, Employment and Migration Working Papers, No. 224, OECD Publishing (Paris 2019). <http://dx.doi.org/10.1787/b5e5f16e-en>.
- DG Employment, Social Affairs and Inclusion, Identification of latest trends and current developments in methods to profile jobseekers in European Public Employment Services: Final Report (Brussels 2015).
- https://webcache.googleusercontent.com/search?q=cache:TuUf_sblU4J:https://www.euro.centre.org/downloads/detail/3167+&cd=1&hl=de&ct=clnk&gl=at.
- DG Employment, Social Affairs and Inclusion, Tackling Long-Term Unemployment through Risk Profiling and Outreach, A Discussion Paper from the Employment Thematic Network, Technical Dossier no 6, Brussels (May 2018).
- <https://www.euro.centre.org/downloads/detail/3167>.
- Dornmayr, H., Projektive Analyse der individuellen Dauer von Arbeitslosigkeit, Forschungsbericht des IBE, Teil I: Textband + Teil II: Datenband (Linz 1996).
- Eberts, R. W., O'Leary, Ch. J., Wander, St. A., Targeting Employment Services, W. E. Upjohn Institute for Employment Research, Klamazoo (Michigan 2002).
- Eco-Austria, Befunde zur Effektivität der Aktiven Arbeitsmarktpolitik, Endbericht im Auftrag des Bundesministeriums für Finanzen (Wien, April 2016).
- Eppel, R., Mahringer, H., Weber, A., Zulehner, Ch., Evaluierung der Eingliederungsbeihilfe, Studie im Auftrag des BMASGK (Wien 2011). http://www.forschungsnetzwerk.at/downloadpub/2011_eingliederungsbeihilfe_bmask_bericht_mai2011.pdf.
- Eppel, R., Horvath, Th., Lackner, M., Mahringer, H., Hausegger, T., Hager, I., Reidl, Ch., Reiter, A., Fridl-Schafferhans, M., Evaluierung von Sozialen Unternehmen im Kontext neuer Herausforderungen, Studie im Auftrag des BMASGK (Wien, April 2014).
- https://www.wifo.ac.at/pubma-datensaetze?detail-view=yes&publikation_id=50691.
- Eppel, R., Leoni, Th., Mahringer, H., Hausegger, T., Reidl, Ch., Weber, F., Einsatz und Wirkung aktiver ar-beitsmarktpolitischer Maßnahmen für Personen mit gesundheitlichen Einschränkungen. Eine Evaluierung für Oberösterreich, Studie im Auftrag des BMASGK (Wien März 2017). [http://www.forschungsnetzwerk.at/downloadpub/s_2017_arbeitsmarktpolitik_einschraenkungen_59356\\$.pdf](http://www.forschungsnetzwerk.at/downloadpub/s_2017_arbeitsmarktpolitik_einschraenkungen_59356$.pdf).
- Frölich, M., Lechner, M., Steiger, H., Statistically Assisted Programme Selection – International Experiences and Potential Benefits for Switzerland, in: Swiss Journal of Economics and Statistics, 139, (May 2003), 311-331.
- https://www.alexandria.unisg.ch/15873/1/Statistically_Assisted_Programme_Selection_International_Experiences_and_Potential_Benefits_for_Switzerland.pdf.
- Gibbons, S., Challenges with Integrating Data-Driven Approaches to Reemployment Services Delivery, Präsentation anlässlich des OECD Technical Workshops „Profiling Tools and their use in active labour market policies“ (Paris, 21 June 2018).
- <http://www.oecd.org/employment/emp/profiling-workshop.htm>.
- Ismail, K., AI vs. Algorithms: What's the difference, CMS-Wire (Oct. 26, 2018).
- <https://www.cmswire.com/information-management/ai-vs-algorithms-whats-the-difference/>.
- Klewais, E., Next Steps @ VDAB. Key challenges for implementation profiling, Präsentation anlässlich des OECD Technical Workshops „Profiling Tools and their use in active labour market policies“ (Paris, 21 June 2018).
- <http://www.oecd.org/employment/emp/profiling-workshop.htm>.
- Kopf, J., AMS-Vorstand Kopf: 'Was die EDV gar nicht abbilden kann, ist die Motivation, in: Der Standard (10. Oktober 2018a). <https://www.derstandard.at/story/2000089096795/ams-vorstand-kopf-menschliche-komponente-wird-entscheidend-bleiben>.

- Ders., Wie Ansicht zur Einsicht werden könnte, Blog vom (14. 11. 2018b).
<https://www.johanneskopf.at/2018/11/14/wie-ansicht-zur-einsicht-werden-koennte/>.
- Ders., Offener Brief an Fr. Prof. Sarah Spiekermann zum Thema Einsatz von KI im AMS, Blog vom (24.9.2019a). <https://www.johanneskopf.at/2019/09/24/offener-brief-fr-prof/>.
- Ders., Ein kritischer Blick auf die AMS-Kritiker, in: Der Standard (25. 09. 2019b).
<https://www.derstandard.de/story/2000109032448/ein-kritischer-blick-auf-die-ams-kritiker>.
- Ders., Der Beipackzettel zum AMS-Algorithmus, in: futurezone (9. 10. 2019c).
<https://futurezone.at/meinung/der-beipackzettel-zum-ams-algorithmus/400641347>.
- Kurekoya, L. M., Review of Profiling Systems, Categorization of Jobseekers and Calculation of Unit Service Costs in Employment Services – Implications and Applications for Slovakia, in: CELSI Research Report No 8 (Bratislava 2014).
- Lopez, P., Reinforcing Intersectional Inequality via the AMS Algorithm in Austria, in: Critical Issues in Science, Technology and Society Studies, Conference Proceedings of the 18th Conference in Graz 2019 (6th-7th May 2019), Graz: Verlag Technische Universität Graz.
- Lopez, P., Interview: <https://netzpolitik.org/2019/mal-sehen-was-der-computer-sagt/>.
- Langenbacher, K., Prognosemodelle zur Ermittlung der Arbeitsmarktchancen von Arbeitslosen, Vortrag, AMS-Seminar (Wien, 18. Jänner 2019).
- Lechner, M., Miquel, R., Werner, St., Wiehler, St., Mikroökonomische Evaluierung der Instrumente der Aktiven Arbeitsmarktpolitik in Österreich, Studie im Auftrag des BMWA (Wien, 2007). http://www.forschungsnetzwerk.at/downloadpub/Mikrooekonomische_Evaluierung_der_Instrumente_der_Aktiven_Arbeitsmarktpolitik_oesterreich.pdf.
- Liebeswar, C., Taschwer, M., Egger-Subotitsch, A., Evaluierung der PPC-Pilotierung, Perspektivencheck, Begleitforschung zur Pilotierung, gekürzter Endbericht, abif (Wien, 1.2.2019). http://www.forschungsnetzwerk.at/downloadpub/PPC_Endbericht_abif.pdf.
- Lischka, K., Klingel, A., Wenn Maschinen Menschen bewerten. Internationale Fallbeispiele für Prozesse algorithmischer Entscheidungsfindung, Bertelsmann Stiftung (Mai 2017).
https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/ADM_Fallstudien.pdf.
- Loxha, A., Morgandi, M., Profiling the Unemployed. A Review of OECD Experiences and Implications for Emerging Countries (2014).
<http://documents.worldbank.org/curated/en/678701468149695960/Profiling-the-unemployed-a-review-of-OECD-experiences-and-implications-for-emerging-economics>.
- Lutz, H., Mahringer, H., Pöschl, A., Evaluierung der österreichischen Arbeitsmarktförderung 2000-2003, Studie im Auftrag des BMASK (Wien 2005).
https://www.wifo.ac.at/jart/prj3/wifo/main.jart?content-id=1454619331110&publikation_id=28498&detail-view=yes.
- Niklas, J., Sztandar-Sztanderska, K., Szymielewicz, K., Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making, in: Fundacja Panoptikon (Warszawa 2015).
https://webcache.googleusercontent.com/search?q=cache:vR6U4G5AP7MJ:https://panoptikon.org/sites/default/files/leadimage-biblioteka/panoptikon_profiling_report_final.pdf+&cd=1&hl=de&ct=clnk&gl=at.
- OECD, Early Identification of Jobseekers at Risk of Long-Term Unemployment. The Role of Profiling, OECD Proceedings (Paris 1998).
- OECD, Profiling tools for early identification of jobseekers who need extra support, in: Policy Brief on Activation Policies (Paris 2018).
<http://www.oecd.org/els/emp/OECD-Profiling-policy-brief-DEC-18.pdf>.
- PLS Ramboll, Study on Early Identification of High Risk Unemployed, Final Report (2001).
www.pls-ramboll.com/homepage/uk/publications/.

- Spiekermann, S., Warum das AMS keine KI auf österreichische Bürger loslassen sollten, in: der Standard (23.9.2019). <https://www.derstandard.at/story/2000108890110/warum-das-ams-keine-ki-auf-oesterreichische-buerger-loslassen-sollte>.
- Dies., Ist der Glaube an zeitsparende AMS-Algorithmen naiv? in: Der Standard (27.9.2019). <https://www.derstandard.at/story/2000109166506/ist-der-glaube-an-zeitsparende-ams-algorithmen-naiv>.
- Synthesis Forschung Gesellschaft m.b.H., Profiling. Grundkonzept, internationale Erfahrungen und Pilotanalysen für das Arbeitsmarktservice (Wien 2007).
- Dies., Integrationschancen. Entwicklung eines Prognosemodells zur KundInnensegmentierung des AMS (Wien, Mai 2016a).
- Dies., Arbeitsmarktchancen. Entwicklung eines Prognosemodells zur KundInnensegmentierung des AMS (Wien, Oktober 2016b).
- Dies., AMS-Chancen 2018. Das AMS-Arbeitsmarktchancen-Modell mit zeitnahe Schätzverfahren und erweiterten Meilensteinen (Wien, Dezember 2017).
- Dies., Das AMS-Arbeitsmarktchancen-Modell. Dokumentation der Methode (Wien, Oktober 2018). http://www.forschungsnetzwerk.at/downloadpub/arbeitsmarktchancen_methode_%20dokumentation.pdf.
- Dies., AMS-Chancen 2019. Das AMS-Arbeitsmarktchancen-Modell (Wien, Jänner 2019a).
- Dies., Personenbezogene Wahrscheinlichkeitsaussagen („Algorithmen“). Stichworte zur Sozialverträglichkeit. Begleitband zur Dokumentation des „AMS-Chancen-Modells“ (Wien, 9. Mai 2019b). https://www.synthesis.co.at/images/PDF/Personenbezogene_Wahrscheinlichkeitsaussagen_Algorithmen_Mai2019.pdf.
- Schweighofer, J., Erzielen die Programme der aktiven Arbeitsmarktpolitik in Österreich ihre beabsichtigten Wirkungen? In: Materialien zu Wirtschaft und Gesellschaft (Wien, März 2013). <https://emedien.arbeiterkammer.at/viewer/ppnresolver?id=AC10765011>.
- Szigetvari, A., Etwas weniger Zwang: AMS fixiert neue Spielregeln für Arbeitslose, in: Der Standard (10.12.2019). <https://www.derstandard.at/story/2000112068281/etwas-weniger-zwang-ams-fixiert-neue-spielregeln-fuer-arbeitslose>.
- Ders., AMS-Algorithmus: Forscher warnen vor Diskriminierung und bemängeln fehlende Transparenz, in: Der Standard (25.2.2020). <https://www.derstandard.at/story/2000114974300/ams-algorithmus-forscher-warnen-vor-diskriminierung-und-bemaengeln-fehlende-transparenz>.
- Weber, F., Hager, I., Krüse, T., Reidl, Ch., Evaluierung des Betreuungsformates für Personen mit multiplen Vermittlungshindernissen (BBEN), Endbericht, AMS (Wien, Jänner 2019). http://www.forschungsnetzwerk.at/downloadpub/2019_BBEN_BBEN-ams_final.pdf.
- Weinberger, F., BBEN: Auswertungen/Evaluierung, Förderausschuss (Wien, 27.2.2019).
- Wijnhoven, M. A., Havinga, H., The Work Profiler: A digital instrument for selection and diagnosis of the unemployed, in: Local Economy, Vol 29 (6-7) (2014) 740-749. https://www.researchgate.net/publication/275458502_The_Work_Profiler_A_digital_instrument_for_selection_and_diagnosis_of_the_unemployed.
- Wimmer, B., Interview mit AMS-Chef Kopf, „Mitarbeiter schätzen Jobchancen pessimistischer ein als der Algorithmus“, in: Futurezone (12.10.2018). <https://futurezone.at/netzpolitik/ams-chef-mitarbeiter-schaetzen-jobchancen-pessimistischer-ein-als-der-algorithmus/400143839>.
- Zweig, K., Algorithmische Entscheidungen: Transparenz und Kontrolle, in: Analysen und Argumente Nr. 338, Konrad Adenauer Stiftung (Januar 2019a). <https://www.kas.de/documents/252038/4521287/AA338+Algorithmische+Entscheidungen.pdf/533ef913-e567-987d-54c3-1906395cdb81?version=1.0&t=1548228380797>.
- Dies., In algorithmischer Sippenhaft, in: Gehirn&Geist 62, (12_2019b).
- Dies., Wo Maschinen irren können. Verantwortlichkeiten und Fehlerquellen in Prozessen algorithmischer Entscheidungsfindung, Bertelsmann-Stiftung (Februar 2018).

<https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WoMaschinenIrrenKoennen.pdf>.

Zweig, K., Krafft, T., Fairness und Qualität algorithmischer Entscheidungen, in: Mohabbat Kar, B., Thapa, E. P., Parycek, P., (Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft, Berlin, Fraunhofer-Institut für offenen Kommunikationssysteme FOKUS (Berlin 2018). <https://www.oeffentliche-it.de/documents/10181/76866/8+Zweig+%26+Krafft+-+Fairness+und+Qualität+algorithmischer+Entscheidungen.pdf>.

Zusammenfassung

Der Text behandelt die Möglichkeiten und Grenzen von *Profiling* in der Arbeitsmarktpolitik: Ausgehend von den Erfahrungen in den USA seit den 1990er-Jahren wird gezeigt, dass beim *Profiling* viele Probleme seit Beginn ungelöst blieben, wie etwa die fehlende Schätzgenauigkeit, der geringe Nutzen von *Profiling*-Ergebnissen für die Zuweisung zu Programmen, die Widerstände von BeraterInnen etc. Damit bleibt die zentrale Erkenntnis für die österreichische Arbeitsmarktpolitik, dass *Profiling* für das *Targeting* beim AMS wenig hilfreich ist. Über die genannten Problem hinaus muss festgehalten werden, dass mit der intransparenten Vorgangsweise des AMS und mit der Wahl des Unternehmens, das *Profiling* im AMS entwickeln und implementieren soll, der Sache des *Profiling* kein guter Dienst erwiesen wurde. Daraus folgt, dass die österreichische Arbeitsmarktpolitik besser heute als morgen den unglücklichen Versuch der Implementierung von *Profiling* beenden sollte.

Abstract

The text deals with the possibilities and limitations of profiling in labour market policy: Based on the experiences in the USA since the 1990s, it is shown that many problems in profiling have remained unsolved since the beginning, such as the lack of estimation accuracy, the low utility of profiling results for allocation to programmes, the resistance of counsellors, etc. Thus the central finding for Austrian labour market policy remains that profiling is of little use for targeting at the AMS. In addition to the problems mentioned above, it must be noted that the non-transparent approach of the AMS and the choice of the company that is to develop and implement profiling in the AMS have not done a good service to the cause of profiling. It follows that Austrian labour market policy should, sooner rather than later, put an end to the unfortunate attempt to implement profiling.

Schlüsselbegriffe: *Profiling*, *Targeting*, Algorithmen, künstliche Intelligenz, Strategiewechsel, Arbeitsmarktpolitik.

Key words: profiling, targeting, algorithms, artificial intelligence, change of strategy, labour market policy.

JEL codes: J64, J65, J68.