

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

*Kerstin Bachberger-Strolz*

## **INHALT**

- 1. Einleitung**
- 2. Wofür eignet sich Profiling, wofür nicht?**
  - 2.1. Die Anfänge der Diskussion in den 1990er Jahren
  - 2.2. In den 2010er Jahren
  - 2.3. Zusammenfassung der internationalen Erfahrungen mit Profiling
  - 2.4. Die Umsetzung von Profiling im AMS durch Synthesis Forschung Ges.m.b.H
    - 2.4.1. Kritik am AMS-Profiling-Ansatz
- 3. Strategiewechsel in der österreichischen Arbeitsmarktpolitik: Es geht um Targeting und nicht um Profiling**
  - 3.1. Wofür möchte das AMS Profiling verwenden?
  - 3.2. Effektivität und Effizienz in der österreichischen Arbeitsmarktpolitik: Bisherige Evaluierungsergebnisse für besondere Problemgruppen
    - 3.2.1. Die grundlegende Problemstellung
    - 3.2.2. Beratungs- und Betreuungseinrichtungen (BBEN)
    - 3.2.3. Frühere Evaluierungsergebnisse zur Wirksamkeit von Programmen für besonders schwierige Problemgruppen am Arbeitsmarkt
  - 3.3. Die Erfahrungen mit „Statistical Assistance for Programme Selection“ (SAPS) in der Schweiz
- 4. Künstliche Intelligenz in der Arbeitsmarktpolitik: Transparenz, Rechenschaftspflicht, Ethik und Menschenrechte.**
  - 4.1. Algorithmen, algorithmische Entscheidungssysteme und künstliche Intelligenz
    - 4.1.1. Künstliche Intelligenz in der Arbeitsmarktpolitik
  - 4.2. Ethische Fragestellungen und Menschenrechte
- 5. Schlussfolgerungen**
- 6. Literaturverzeichnis**
- 7. Anmerkungen**

## 1. Einleitung

„Alles ein großer Pallawatsch“, möchte man feststellen angesichts der bisherigen Diskussionen zum so genannten „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 der AMS-KundInnen“, „Effektivität und Effizienz in der Arbeitsmarktpolitik“, „Sozialpolitische Verantwortung des AMS“, „Profiling“, „Algorithmus“, „Künstliche Intelligenz“, etc. kunterbunt und, leider oft auch ausschließlich interessensgetrieben, heftig durcheinandergewirbelt wurden. Besonders symptomatisch war die Verwechslung der Begriffe Profiling<sup>1</sup> und Algorithmus<sup>2</sup> bzw. was mit den jeweiligen Begriffen im Einzelnen gemeint ist. Dies führte zu zwei folgenschweren Irrtümern, von denen sich die Diskussionen bis dato nicht „erholt“ haben bzw. die noch immer nicht ausgeräumt werden konnten:

- (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.<sup>3</sup>
- (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.

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 dann tatsächlich gemacht werden wird. In der Realität wird sich wahrscheinlich sehr wenig ändern (jedenfalls für weniger als 10% aller Arbeitslosen), weil die Zielgruppenlogik bei den Programmen nicht geändert wird. Damit stellt sich aber die Frage: Wozu dann überhaupt dieser große Aufwand? Bei genauerem Hinsehen muss eine Antwort ausbleiben.

Bisher überhaupt außer Acht gelassen wurde darüber hinaus die Tatsache, dass in anderen Ländern Profiling als ein „early warning system“ im Falle von schlechten Beschäftigungsaussichten („Hier besteht besonderer Handlungsbedarf!“) verwendet wurde und wird. In diesem Falle stehen die Personen mit schlechten Jobaussichten 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 sollte; 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.<sup>4</sup> All diese Versäumnisse lassen sich durch private Blog-Beiträge von AMS-Vorständen jedenfalls nicht aus der Welt schaffen, insbesondere dann nicht, wenn sich diese Beiträge über eine affirmative, beschönigende und propagandahafte Beschreibung der Sachverhalte nicht hinausbewegen.

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 Diskussionen<sup>5</sup>**

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“<sup>6</sup>, 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 long-standing and basic endeavor 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.“<sup>7</sup>

Im Jahr 2004 hat man in Schweden im Rahmen einer internationalen Konferenz („Can statistical models increase the efficiency at the employment office?“, Uppsala, December 3, 2004) überlegt, statistische 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 „decision support“ dienen, durch die effiziente Verwendung der vorhandenen Informationen und durch bessere Selektion; 2) die Effizienz erhöhen, d.h. die Nettoerträge (Erträge minus Aufwendungen) der Maßnahmen maximieren und 3) die Chancengleichheit erhöhen. Die gesamten 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 („equity goal“) gefördert werden sollen. Nur wenn die Heterogenität der Maßnahmenwirkungen derart ist, dass beispielsweise die größten Wirkungen bei den Personengruppen auftreten, die die höchste Wahrscheinlichkeit aufweisen, langzeitarbeitslos zu werden, fallen die beiden Zielsetzungen zusammen. Bei homogenen Maßnahmenwirkungen und homogenen Grenzkosten können grundsätzlich keine Effizienzvorteile lukriert werden. In Berger et. al. (2001, 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“<sup>8</sup>. 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.

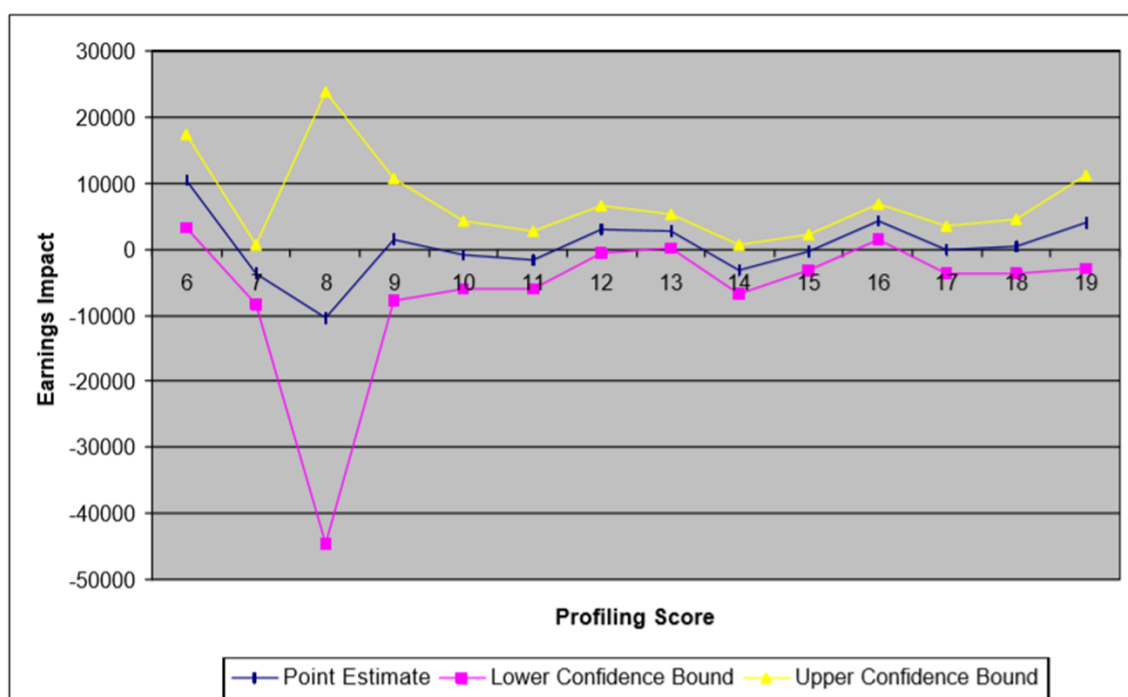
Berger et. al. (2001, 59) fassen ihre Ergebnisse wie folgt zusammen:

„First, we demonstrate that it is possible to effectively predict the duration of UI spells<sup>9</sup> ... 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 service<sup>10</sup> 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, 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. They also suggest the value of further thought and study before ex-tending profiling to other programs.”<sup>11</sup>

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



Quelle: Berger et.al. (2001) 81

Die vorangegangene Abbildung 1 veranschaulicht den, in den vorangegangenen Zitaten erwähnten, fehlenden Zusammenhang noch einmal grafisch: Zwischen der Länge der prognostizierten Dauer der Arbeitslosigkeit (Profiling-Variable) und dem Ergebnis des „Worker Profiling and Reemployment Service (WPRS)“ (hier der Einkommenseffekt) besteht kein(!) Zusammenhang. Im Effekt hilft also Profiling nicht, die Effizienz von Maßnahmen (WPRS) zu erhöhen, weil es entlang der prognostizierten Arbeitslosigkeitsdauer keine Maßnahmenheterogenität gibt, 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 versuchen Bryson/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<sup>12</sup>. Es wurden drei Ergebnis-Variablen<sup>13</sup> für die drei Hauptgruppen von LeistungsbezieherInnen „Kranke/Behinderte“, „Allein erziehende Eltern“ und „ALG-BezieherInnen“ prognostiziert. 12 Pilot-Gebiete wurden 12 zufällig ausgewählten Vergleichsregionen gegenübergestellt. Die zentrale Frage der ausreichenden Genauigkeit der Prognosen - die Beantwortung dieser Frage kann lt. AutorInnen letztlich nur von der Politik vorgenommen werden - wurde über zwei Verfahrensschritte angegangen: a) den Vergleich der Durchschnitte einzelner Quintile im Vergleichssampel und b) durch die Festlegung einer Entscheidungsregel (30%, 50% oder 70% Maßnahmenteilnehmer) und die anschließende Ermittlung von „Korrektter Teilnahme“ (das Modell prognostiziert im Vergleichssampel - 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“ (siehe nachfolgende Tabelle 1). Eine zentrale Annahme in dieser Machbarkeitsstudie besteht darin, davon auszugehen, dass Personen mit der höchsten Wahrscheinlichkeit, langzeitarbeitslos zu werden, die höchsten Maßnahmenwirkungen erzielen.

Ergebnisse der Untersuchung:

- Profiling ist besser als eine Zufallszuteilung zu Maßnahmen - eine vergleichsweise niedrige Hürde! Ein Vergleich mit der Maßnahmenzuteilung von Beratern in Großbritannien konnte aus Datengründen nicht erfolgen. „Wrong denial“ und „wrong treatment“ bleiben jedoch ein gravierendes Problem (im Ausmaß von 26-44% aller Personen)!
- Die Genauigkeit der Prognose hängt 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 Schätzmethode zeigen keinen großen Performance-Unterschiede.
- Modelle mit mehr erklärenden Variablen prognostizieren meist etwas genauer.
- Die Unterschiede in verschiedenen Bereichen der Verteilung „mean predicted outcome“ liegen etwa zwischen 35%-50%, die Prognosegüte ist jedoch offensichtlich nicht abhängig davon, welcher Bereich der Verteilung betrachtet wird.

- Zur nachfolgenden Tabelle 1: 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.

**Tabelle 1: Erfolg bei der Zielgruppenorientierung von Maßnahmenteilnahmen**  
(Logit-Schätzer, Profiling Variable: Keine Beschäftigung)

	Correctly treated	Correctly denied	Wrongly treated	Wrongly denied
<b>70% treated</b>				
Pars	55	16	15	14
Full	57	17	14	12
<b>50% treated</b>				
Pars	43	22	9	26
Full	45	24	7	24
<b>30% treated</b>				
Pars	29	37	3	41
Full	30	29	2	39

Quelle: Bryson/Kasparova (2003)

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

- Kleinst-Quadraten-Schätzer 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).
- Überraschenderweise verbessern die Variablen „lokalen Arbeitslosenquoten“ und das „Beschäftigungswachstum in der Industrie“ (als Konjunkturvariablen) die prognostischen Eigenschaften der Modelle nicht.
- Modelle mit mehr erklärenden Variablen verbessern ebenfalls 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. In den 2010er Jahren<sup>14</sup>

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 11 Länder in der OECD statistisches Profiling, was aber im Umkehrschluss heißt, dass 25 Staaten die Anwendung dieses Tools nicht für sinnvoll erachten (siehe die nachfolgende Tabelle 2). Nach etwa 25 Jahren des Einsatzes von Profiling, welche Erfahrungen wurden mit diesem Instrument 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 deadweight 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.“<sup>15</sup>

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 „click data“ 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).<sup>16</sup>

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 (siehe Tabelle 2), verbessert werden kann. Die Präzision der Schätzungen hängt offensichtlich von der Qualität und Art der Input-Daten ab, aber nicht von der Wahl des Modelles und auch nicht von der Verwendung von Verhaltensdaten.<sup>17</sup> 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 dieses Instrumentes fundamental in Frage gestellt.<sup>18</sup> 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.<sup>19</sup>

Langenbacher (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 Seite der BeraterInnen („ensuring buy-in by case workers“, siehe Zitat am Beginn von Kapitel 2.2.). 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)”<sup>20</sup>. 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 geringgehalten werden sollen, dann ist klar, dass mit „job-ready“ 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.<sup>21</sup>
- Schließlich wird im Zitat oben (Desiere et.al. 2019) etwas verstohlen („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 des Profiling genau gemacht werden soll bzw. wie diese verwendet werden können, um die Ergebnisse der Programmteilnahmen zu verbessern (Stichwort: Effektivität und Effizienz).

uf die Frage, was man angesichts der offensichtlichen Grenzen des Einsatzes von Profiling tun könnte, stellt Langenbacher (2019) fest, dass die Daten regelmäßig aktualisiert werden müssten; umfangreichere Datensätze (die auch „soft skills“ und Suchverhalten bzw. Motivation umfassen sollten) 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.

Tabelle 2: Charakteristika von statistischen Profiling Modellen in der OECD

	Outcome (probability of)	Data derived from	Type of data sources (in addition to socioeconomic info)					Statistical model	Accuracy	Compulsory/voluntary use by	
			Labour market history <sup>a</sup>	Hard skills	Soft skills	Motivation Jobseekers' behaviour	Opportunities Regional labour market info			Jobseekers	Caseworkers
Australia	Long-term unemployed (12 months)	Personal interview; online trial ongoing	Yes	Yes	No	No	No	Logistic regression		Compulsory	Compulsory
Austria	Labour market integration probability <sup>b</sup>	Administrative data	Yes	Yes	No	No	Yes	Logistic regression	80%-85%	Compulsory	
Belgium (Flanders)	Long-term (>6 months) unemployed	Administrative data; "click" data	Yes	Yes	No	Yes	No	Random forest model	67% (AUC ~0.76)	Compulsory	Compulsory
Denmark	Long-term (>26 weeks) unemployed	Online questionnaire; Administrative data	Yes	Yes	Yes	Yes	No	Big data model	>60%	Voluntary	Voluntary
Ireland	Probability of exit to employment within 12 months	Questionnaire as part of benefit claim process, administrative data	Yes	Yes	No	Yes	Yes	Probit regression	70% - 86%	Compulsory	Compulsory
Italy	Long-term unemployed (12 months)	Administrative data	Yes	Yes	No	No	Yes	Logistic regression		Compulsory	Compulsory
Latvia	Long term unemployed (12 month)	Personal (individual) interview, questionnaire, administrative data	Yes	Yes	Yes	Yes	Yes	Factor analysis	No data yet	Compulsory at PES & voluntary online	Compulsory (advisory for now)
Netherlands	Long-term unemployed (12 months)	Online questionnaire		Yes	Yes	Yes		Logistic regression	70%	Voluntary	Compulsory
New Zealand	Lifetime income support costs (LET), change in lifetime income support and staff costs from receiving a case management service (SEM)	SEM/LET are based on administrative data	Yes	Yes	No	No	No	Random forest (LET), Gradient boosting (SEM)	AUC: 0.63 - 0.83	Compulsory for jobseekers; opt-in for other PES clients	Compulsory
Sweden	LTU (6 months)	Administrative data	Yes	Yes	No	No	Yes	Logistic regression			Voluntary
US	Exhausting the 26-week entitlement to UI benefits	Online questionnaire; Administrative data		Yes	No	No	Yes	Logistic regression		Compulsory	Compulsory

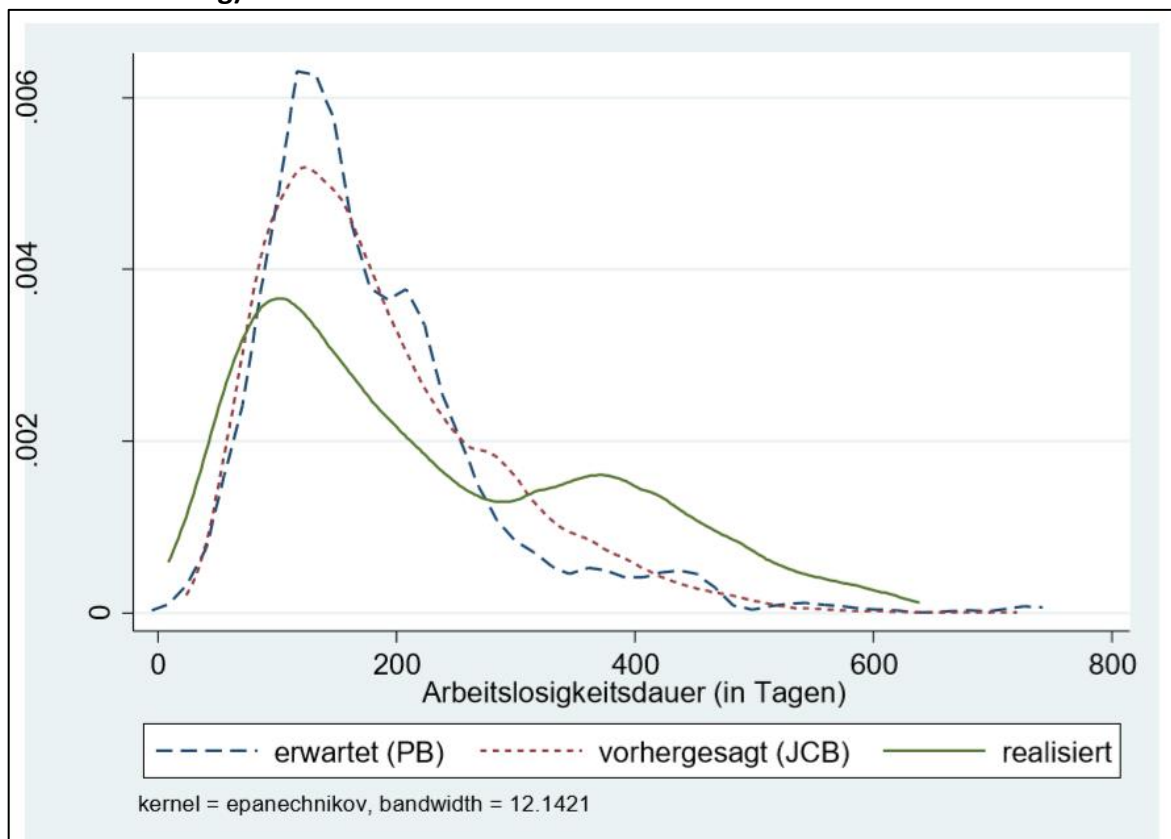
*Notes:* a. Input data includes information on prior periods of employment, unemployment, inactivity; occupation(s) held; sector(s) worked in.  
b. Outcome variable in Austria: Labour market integration probability measured in short and long term (short term: 3 months of unsubsidised employment within 7 months; long term: 6 months of unsubsidised employment within 24 months).  
*Source:* Authors' compilation.

Quelle: Desire et. al. (2019) 11

Abschließend zu diesem Unterkapitel sei noch auf drei interessante internationale Beispiele hingewiesen:

(1) Zunächst auf einen Vergleich der Prognose von Arbeitslosigkeitsdauern in der Schweiz: Wer kann diese genauer prognostizieren, Profiling Modelle oder BeraterInnen? Und wie sieht ein Vergleich mit den tatsächlichen Dauern aus?<sup>22</sup>

**Abbildung 2: Vergleich der Verteilungen der erwarteten (Berater), vorhergesagten (Job-Chancen-Barometer-Profiling) und realisierten Arbeitslosigkeitsdauern (Pilotversuch Schweiz-Freiburg)**



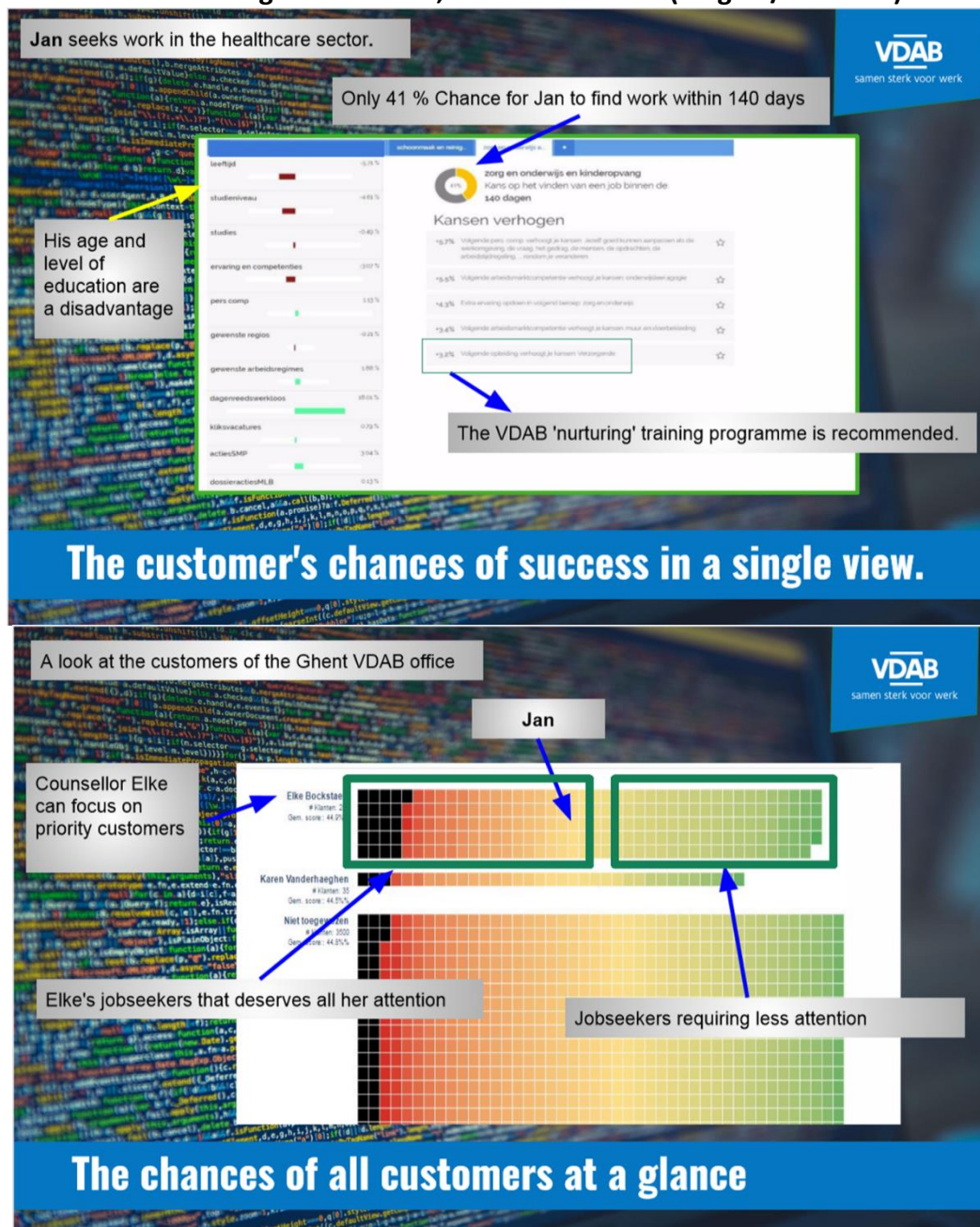
Quelle: Arni/Schiprowski (2015)

Wie in der Abbildung 2 zu sehen ist, haben die BeraterInnen bei der Prognose der Arbeitslosigkeits-Dauern einen Bias hin zur Mitte der Verteilung, d.h. sie unterschätzen die Dauern deutlich, weil sie zu wenige lange Dauern annehmen. Der Job-Chancen-Barometer (Profiling) unterliegt einem ähnlichen Schätzfehler, aber in einem geringeren Ausmaß, d.h. mit Profiling lassen sich die Dauern etwas genauer vorhersagen.

(2) In Belgien/Flandern wird eine Form von Profiling verwendet, das mit Elementen von Künstlicher Intelligenz verbunden ist (siehe Abbildung 3)<sup>23</sup>: Am Beispiel einer Person mit dem Namen „Jan“, der vor Arbeitslosigkeit im Gesundheitsbereich gearbeitet hat und der für den Zeitraum der nächsten 140 Tagen mit einer Wahrscheinlichkeit von 41% wieder einen Job finden wird, wird gezeigt, mit welchen Informationen, die auf der Maske im PC der Beraterin „Elke“ aufscheinen, gearbeitet werden kann. Auf Grund des Profiling wird einerseits gezeigt, dass das Alter und das Ausbildungsniveau von „Jan“ als Nachteil für die Reintegration in den Arbeitsmarkt gesehen werden und

andererseits, welcher Kurs für den Arbeitslosen in Frage käme, weil ähnliche Personen wie Jan davon bereits in der Vergangenheit profitiert haben.

**Abbildung 3: Maske für die Beraterin „Elke“, die den Arbeitslosen „Jan“, der im Gesundheitsbereich gearbeitet hat, unterstützen soll (Belgien/Flandern)**



Quelle: Klewais (2018)

(3) In den Niederlanden wiederum werden, im Rahmen einer ersten Diagnose, ebenfalls der/dem BeraterIn über das Profiling die Chance auf Wiedereingliederung in den Arbeitsmarkt (im vorliegenden Fall beträgt diese 83%) angezeigt, wie auch die förderlichen und die hinderlichen Eigenschaften der entsprechenden, arbeitslosen Person. Interessant dabei ist die Vielzahl an „soft factors“, die (neben den „hard factors“) über einen Fragebogen von 20 Fragen ermittelt werden. Diese 20 Fragen sind Resultat jener (im niederländischen Fall 11) Faktoren, die für die Vorhersage der Reintegrationswahrscheinlichkeiten die größte Relevanz haben.<sup>24</sup>

Abbildung 4: Maske für BeraterInnen bei der niederländischen Arbeitsvermittlung

Profile   <span>Menu ▼</span>		Chance at work resumption	
		83%	
Factor	Diagnosis	Indication	
Age		■	
Years employed in last job		■	
Problems understanding Dutch		■	
Views on return to work		■	
Feeling too ill to work		■	
Job search behaviour: contact with employers		■	
Job search intention	Hindering	■	
External variable attribution		■	
General work ability		■	
Physical work ability		■	
> Mental work ability	Hindering	■	

Quelle: Wijnhoven und Havinga (2014)

### 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 (Schätzgenauigkeit, Profiling-Ergebnisse nutzen für die Zuweisung zu Programmen, BeraterInnen-Widerstände, etc.) 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 durch Synthesis Forschung Ges.m.b.H: Eine große Blackbox!

Abgesehen von Vorläufern<sup>25</sup>, hat die aktuelle Debatte zum Profiling im AMS etwa im Jahr 2016 begonnen.<sup>26</sup> Die Synthesis-Forschung Ges.m.b.H. wurde beauftragt, ein „Integrationschancen-Prognosemodell zur Kundensegmentierung im AMS“ zu entwickeln. In einer sehr knappen Darstellung auf nicht einmal 10 spärlich beschriebenen Seiten wurde die methodische Vorgangsweise – für die breitere Öffentlichkeit zugänglich – dokumentiert.<sup>27</sup> Demnach hat Synthesis ein „Modell zur Prognose der regionalspezifischen Arbeitsmarkt-Integrationschancen von vorgemerkten Arbeitslosen“<sup>28</sup> mittels logistischer Regressionen geschätzt, wobei das Modell persönliche Merk-

male (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 Maßnahmeninsatz und Typ des regionalen Arbeitsmarktgeschehen) berücksichtigt. Von ca. 1,2 Mio. AMS-Geschäftsfällen im Jahr konnten im Basismodell<sup>29</sup> nur für ca. 440.000 Fälle „voll valide Schätzungen“<sup>30</sup> durchgeführt werden, weil nur für diese Personen im Zeitraum von 4 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.<sup>31</sup>

**Abbildung 5: Schätzergebnisse, logistische Regression, Basis-Population**

$$\begin{aligned}
 & \text{BE\_INT} \\
 & = f ( 0,10 \\
 & \quad - 0,14 \times \text{GESCHLECHT\_WEIBLICH} \\
 & \quad - 0,13 \times \text{ALTERSGRUPPE\_30\_49} \\
 & \quad - 0,70 \times \text{ALTERSGRUPPE\_50\_PLUS} \\
 & \quad + 0,16 \times \text{STAATENGRUPPE\_EU} \\
 & \quad - 0,05 \times \text{STAATENGRUPPE\_DRITT} \\
 & \quad + 0,28 \times \text{AUSBILDUNG\_LEHRE} \\
 & \quad + 0,01 \times \text{AUSBILDUNG\_MATURA\_PLUS} \\
 & \quad - 0,15 \times \text{BETREUUNGSPFLICHTIG} \\
 & \quad - 0,34 \times \text{RGS\_TYP\_2} \\
 & \quad - 0,18 \times \text{RGS\_TYP\_3} \\
 & \quad - 0,83 \times \text{RGS\_TYP\_4} \\
 & \quad - 0,82 \times \text{RGS\_TYP\_5} \\
 & \quad - 0,67 \times \text{BEEINTRÄCHTIGT} \\
 & \quad + 0,17 \times \text{BERUFSGRUPPE\_PRODUKTION} \\
 & \quad - 0,74 \times \text{BESCHÄFTIGUNGSTAGE\_WENIG} \\
 & \quad + 0,65 \times \text{FREQUENZ\_GESCHÄFTSFALL\_1} \\
 & \quad + 1,19 \times \text{FREQUENZ\_GESCHÄFTSFALL\_2} \\
 & \quad + 1,98 \times \text{FREQUENZ\_GESCHÄFTSFALL\_3\_PLUS} \\
 & \quad - 0,80 \times \text{GESCHÄFTSFALL\_LANG} \\
 & \quad - 0,57 \times \text{MN\_TEILNAHME\_1} \\
 & \quad - 0,21 \times \text{MN\_TEILNAHME\_2} \\
 & \quad - 0,43 \times \text{MN\_TEILNAHME\_3}
 \end{aligned}$$

Quelle: Synthesis (2018) 11

In der vorangegangenen Abbildung 5 ist ein Schätzergebnis (von sehr, sehr vielen) abgebildet, das die kurzfristigen Integrationschancen der Basispopulation zeigt: Bezogen auf eine Referenzgruppe<sup>32</sup>, 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 einen Job zu finden. Bleiben alle anderen Merkmale gleich und es ändert sich beispielsweise nur das Geschlecht (der Geschäftsfall besteht aus einer Frau, alles ander gleich), 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 4 Jahren die Wiederbeschäftigungswahrscheinlichkeit derart nach oben treiben sollte (+1,98). Die letzte Variable hat offensichtlich das falsche Vorzeichen, was von Synthesis unkommentiert bleibt.

**Tabelle 3 Geschäftsfälle 30.4.2016 mit Integrationschancenwerten des Modells 2019**

	Zahl der Geschäftsfälle	Median für die jeweilige Zielfunktion		Anteil der Geschäftsfälle im Bereich <sup>1</sup>		
		kurzfristig	langfristig	»C«	»B«	»A«
<b>Gesamte einbezogene Population</b>	<b>446.229</b>	<b>16%</b>	<b>38%</b>	<b>33%</b>	<b>61%</b>	<b>7%</b>
Frauen	201.901	17%	40%	29%	65%	6%
Männer	244.328	15%	36%	36%	57%	7%
<b>Alter</b>						
bis 18 Jahre	18.509	12%	36%	25%	74%	1%
19 bis 29 Jahre	118.678	23%	52%	15%	78%	7%
30 bis 49 Jahre	198.861	18%	42%	28%	65%	7%
50 Jahre u. älter	110.181	6%	17%	62%	32%	6%
<b>Bundesländer</b>						
Burgenland	11.994	16%	38%	32%	63%	5%
Kärnten	30.274	19%	42%	28%	62%	9%
Niederösterreich	69.100	16%	39%	32%	65%	4%
Oberösterreich	51.214	19%	44%	25%	71%	4%
Salzburg	21.136	34%	60%	14%	63%	23%
Steiermark	54.181	19%	42%	27%	66%	6%
Tirol	32.996	43%	67%	13%	57%	29%
Vorarlberg	14.395	25%	50%	20%	67%	13%
Wien	160.939	9%	27%	46%	53%	1%

<sup>1</sup> Die Zuordnung zu den Bereichen (Segmentierung) folgt dem IC-Wert (Arbeitsmarktchance) in den verschiedenen Zielfunktionen:

- Bereich »C«: IC-Wert bei langfristiger Perspektive < 25%
- Bereich »A«: IC-Wert bei kurzfristiger Perspektive > 66%
- Bereich »B«: Restgruppe

Quelle: Synthesis (2019a) 16

Die vorangegangenen Tabelle 3 soll einen Einblick in die Stabilität der Schätzergebnisse von Synthesis in Bezug auf die Zuordnung zu den Gruppen mit niedrigen („C-Bereich“) und mit hohen („A-Bereich“) Reintegrationschancen geben. 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.<sup>33</sup> 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 Arbeitsmarktgruppen, insbesondere zwischen Wien und Tirol.

#### **2.4.1. Kritik am AMS-Profilings-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, der Verletzung von Grundrechten durch die Verwendung von Algorithmen, etc. Nachfolgend wird v.a. auf technische Aspekte eingegangen.

##### **Mangelnde Transparenz**

Auch wenn AMS-Vorstand Kopf nicht müde wird, das Gegenteil zu behaupten<sup>34</sup>: Das AMS-Profilings wurde bisher völlig intransparent entwickelt und implementiert, jegliche Standards (etwa 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.“<sup>35</sup>

Den AutorInnen des Beitrages ist völlig Recht zu geben: Abgesehen von einer(!) öffentlich zugänglichen Publikation<sup>36</sup>, die den methodischen Vorgang nur spärlich dokumentiert, sind keine weiteren Informationen erhältlich: Einmal spricht Herr Kopf von „7500 unterschiedlichen Teilmodellen“<sup>37</sup>, ein anderes Mal von „mehr als 90 verschiedenen Algorithmen“<sup>38</sup> – 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. 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.<sup>39</sup> 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“.<sup>40</sup> 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 gib es leider keine Antworten, weil diese auch auf Nachfragen nicht zur Verfügung gestellt wurden.<sup>41</sup> Anscheinend ist ein Teil dieser Kritik bei Synthesis sogar bereits angekommen, es wird in einem aktuellen Papier davon gesprochen, betroffenen Personen niederschwellig in den Entscheidungsprozess einzubinden, die Profiling-Ergebnisse nur als „zweite Meinung“ zu werten und Transparenz herzustellen („Darüber lohnt es (als vertrauensbildenden Maßnahme gegenüber der Zivilgesellschaft), Transparenz zu schaffen; über den ‚Algorithmus‘ selbst und die Befunde der Technikfolgenabschätzung.“<sup>42</sup> Diese Einsicht kommt spät und sie kommt einseitig, „ex cathetra“ gewissermaßen, von oben herab verkündet!

### **Treffergenauigkeit**

Die Brauchbarkeit 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“).<sup>43</sup> Synthesis gab bekannt, dass für ca. 440.000 Geschäftsfälle (von insgesamt 1,2 Mio.) voll valide Schätzungen möglich waren, die Trefferquote lag zwischen 80-85%.<sup>44</sup> 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 Fußnote 37). 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: Es wird angenommen, sie wären arbeitslos, sie sind aber beschäftigt. Führt man sich diese absoluten Zahlen vor Augen, dann müssen starke Zweifel am Nutzen von Profiling in der Arbeitsmarktpolitik aufkommen.

### **Diskriminierung**

Diskriminierung 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 Fragen der Motivation, der Arbeitsmarktnähe, der 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 caseworker-based profiling, but also rule-based profiling, are also prone to (statistical) discrimination.”<sup>45</sup>

(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 verfügbar. Ob schikanoöses 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“*.<sup>46</sup> 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 werden Menschen sehr ungern zugeben, dass ihr Verhalten diskriminierend ist – weshalb sie diese zu verbergen suchen werden.

Weitere Kritikpunkte am AMS-Profiling-Zugang (Anwendung von Künstlicher Intelligenz ohne sozialwissenschaftliche Begleitung, der fehlenden 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<sup>47</sup>, also die Verbesserung von Effektivität und Effizienz 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 keine 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? Oder anderes gefragt: Sitzen die richtigen Personen in den richtigen Kursen und ist das Design der Kurse derart, dass Effektivität vor allem für „hard to place people“ gewährleistet ist?

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.“<sup>48</sup>

„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.“<sup>49</sup>

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-Vorstandes Kopf als bestürzend bewertet werden: Er stellt doch tatsächlich nach über 50 Jahren Instrumenteneinsatz fest, dass das AMS schauen müsse „welche Fördermaßnahmen wie wirken“ und dass „bestehende Förderinstrument 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.<sup>50</sup> 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 „Hardcore“ Evaluierungsergebnisse dazu, dass BBEN dieselben Effekte generieren wie SÖBs (es gibt sich nicht, siehe unten) oder sind es nur die üblichen „Brutto-Beschäftigungsaufnahmen“<sup>51</sup>. Versucht man die Aussagen ernst zu nehmen, dann kann man sich des Eindruckes 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 eingereicht werden. Auch für Ältere über 50 Jahren 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 Jobangebote 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.<sup>52</sup>
- 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 BeraterIn ist immer dann möglich, wenn „durch die in Aussicht genommene Förderung eine entsprechende Erhöhung der Integrationschancen erwartet“ wird.<sup>53</sup> Eine Herabsetzung der Arbeitsmarktchancen von mittel/hoch auf niedrig kann nur erfolgen, wenn davor ein externer Perspektivencheck durchgeführt wurde (siehe unten).<sup>54</sup>

- 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.<sup>55</sup>

### **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.<sup>56</sup> Denn dadurch ergibt sich das Problem, dass nicht klar ist, was der folgenden 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 auf Maßnahmen hin 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 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 zur Unkenntlichkeit. Was wäre in dieser Situation eine effiziente 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 in 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/Sozialversicherungs-Beiträge 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 eine Vielzahl von 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 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 3, 6 oder 9 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 Kopf auf AMS-Analysen<sup>57</sup> in einem frühen Stadium der Pilotphase: Diese Bruttoergebnisse zeigten Beschäftigungsaufnahmen 3 Monate nach Beendigung der BBEN-Teilnahmen von ca. 33%, was für diese Problemgruppe einen sehr hohen Wert darstellt; allerdings waren bei mehreren Tausenden Teilnahmen 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 5-18% aller AbgängerInnen aus BBEN. Werden etwas spezifischer Problemgruppen mit Mehrfachbelastungen (Alter über 45 Jahre, Behinderung, max. Pflichtschule, AL vorangegangen Arbeitslosigkeitsdauer über 1 Jahr) verglichen, die 2018 entweder in einem SÖB oder in BBEN waren, dann zeigt sich für den Bestandswert 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%.<sup>58</sup> 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 Maßnahnumsetzung, die allerdings keine klassische Wirkungsanalyse darstellt.<sup>59</sup> Es wurden dabei 14 telefonische Interviews mit AMS-LeiterInnen bzw. –BeraterInnen und TrägervertreterInnen, 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.<sup>60</sup> 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<sup>61</sup>, die für 12 Monate in den BBEN verbleiben konnten.
- Und eine Fiskalanalyse des WIFO<sup>62</sup> 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 besonders schwierige Problemgruppen am Arbeitsmarkt**

*State of the Art*-Maßnahmenevaluierungen zeigen ein gemischtes Bild der kausalen Wirkungen von aktiver Arbeitsmarktpolitik in Österreich.<sup>63</sup> Manche Programme wirken nicht für den Durchschnitt, d.h. etwa im Vergleich einer Nachbeobachtungsperiode von 3 Jahren hat die Kontrollgruppe beispielsweise mehr Tage an 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 (für diese sind häufig positive Effekte festzustellen), Jugendliche und Ältere, gering und höher Qualifizierte, InländerInnen und AusländerInnen, etc.

Frühe Evaluierungen des gesamten Förderprogrammes des AMS<sup>64</sup> zeigten etwa, dass sich die durchschnittlichen positiven Beschäftigungswirkungen von Qualifizierungsprogrammen erst am Ende einer drei jährigen Nachbeobachtungsperiode einstellen, bei Kurskostenbeihilfen (Weiterbildungen, die auf Initiative der Arbeitslosen zurückgehen) 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 (hier handelt es sich um eine zeitlich befristete Lohnsubvention) sind etwa 50% Mitnahmeeffekte zu berücksichtigen (in diesen Fällen wäre eine Einstellung des Arbeitslosen auch ohne Förderung erfolgt); diese Förderung wirkt v.a. bei älteren Arbeitslosen und bei Frauen.<sup>65</sup>

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 Untersuchung<sup>66</sup> der Sozialökonomischen Betriebe (SÖB) noch zu dem Ergebnis kam, dass dieses Programm für Niedrigqualifizierte (maximal Pflichtschulabschluss) im Gegensatz zu höher qualifizierten Arbeitslosen durchaus zu positiven Beschäftigungseffekten gegenüber einer Kontrollgruppe führt, kommt eine rezentere Studie<sup>67</sup> 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. Dieses Ergebnis ist dem AMS schon seit mehreren Jahren bekannt, man hat bisher jedoch darauf noch nicht reagiert - etwa in Form von Änderungen in den Förderrichtlinien.

- Für die Gruppe der gesundheitlich beeinträchtigten Arbeitslosen in Oberösterreich haben WIFO/Prospect 2017 eine Untersuchung<sup>68</sup> 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 ersten 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.

In die Richtung der beiden letztgenannten Evaluierungen sollte verstärkt weiter geforscht werden, um die Programmwirkungen der Personen mit multiplen Vermittlungshemmnissen zu verbessern.

### **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 die die „outcomes“ 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.”<sup>69</sup>

Die nachfolgende Tabelle 4 zeigt die Ergebnisse einer simulierten Reallokation der TeilnehmerInnen an Gruppen von arbeitsmarktpolitischen Programmen in der Schweiz: Es wird die aktuelle Verteilung auf die Programme („Actual Allocation“) verglichen mit einer zufälligen und verschiedenen Varianten von optimalen Verteilung – in Bezug auf die Ergebnis-Variablen Beschäftigung und Einkommen. Der obere Teil der Tabelle zeigt die Verteilung der TeilnehmerInnen auf die verschiedenen Programmtypen, der untere Teil zeigt die „outcomes“. Die zentrale Erkenntnis aus dieser Simulation ist, dass die Beschäftigungsanteile 12 Monate nach Programmteilnahme mit Hilfe von SAPS von 49,8% auf 57,7% (Annahme: 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 Kursen und Programmen kann mit Hilfe von statistischen Methoden deutlich gesteigert werden.

Leider hat sich in einem daran anschließenden Experiment in fünf Regionen (50% der zufällig ausgewählten BeraterInnen konnten in Basel, Bern, Genf, St. Gallen und Zürich von Mai bis Dezember 2005 SAPS nützen, die andere Hälfte nicht) 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 Untersuchen

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 diese Tool in der Praxis nur wenig genutzt wurde.<sup>70</sup> Dies ist ein deutliches Indiz dafür wie wichtig es ist, die institutionellen Logiken und möglichen Widerstände frühzeitig zu berücksichtigen.

**Tabelle 4: Ergebnisse für Programmteilnahmen nach simulierter Reallokation der TeilnehmerInnen (Schweiz)**

Allocation (in %)	Actual Allocation	Random Allocation	Allocation to largest estimate	SAPS- Allocation (MCB 50%)	SAPS- Allocation ,without other'	only Non-Participation
<b>Non-Participation</b>	30.3	30.3	16.7	29.1	30.6	100
<b>Personality Courses</b>	4.3	4.3	1.8	2.3	2.3	-
<b>Language Courses</b>	9.0	9.0	11.2	7.5	7.7	-
<b>Computer Courses</b>	7.9	7.9	6.3	6.5	6.8	-
<b>Vocational Training</b>	5.1	5.1	11.6	5.5	5.7	-
<b>Other courses</b>	1.7	1.7	18.7	4.3	-	-
<b>Employment Programme (individual)</b>	7.4	7.4	7.0	5.6	5.6	-
<b>Employment Programme (collective)</b>	7.4	7.4	3.5	4.3	4.4	-
<b>Temporary Wage subsidy</b>	26.9	26.9	23.2	34.8	36.9	-
Employed after 7 months (%)	38.9	38.3	43.8	43.0	43.4	44.3
Employed after 12 months (%)	49.8	49.5	64.5	57.7	56.9	52.1
Employed after 17 months (%)	62.4	61.9	71.1	66.4	65.7	61.6
Income after 7 months	1'280	1'250	1'460	1'410	1'420	1'420
Income after 12 months	1'580	1'560	2'030	1'800	1'780	1'610
Income after 17 months	1'960	1'950	2'300	2'110	2'080	1'870
Costs of ALMP (CHF)	1'500	1'500	1'740	1'160	1'090	-

Simulation of potential outcomes for all 28'130 individuals. Basis specification

Quelle: Fröhlich et.al. (2003)

#### 4. **„Künstliche Intelligenz“ in der Arbeitsmarktpolitik: Transparenz, Rechenschaftspflicht, Ethik und Menschenrechte!**

##### 4.1. Algorithmen, algorithmische Entscheidungssysteme und künstliche Intelligenz

“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’.”<sup>71</sup>

Algorithmen können also als Handlungsanweisungen beschrieben werden, die von der Eingabe (also den Ausgangsinformationen) zu einem Ergebnis bzw. zur Ausgabe mit gewünschten Eigenschaften führen.<sup>72</sup> Ein Algorithmus kann etwa an Hand 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 entscheidet. 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.“<sup>73</sup>

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

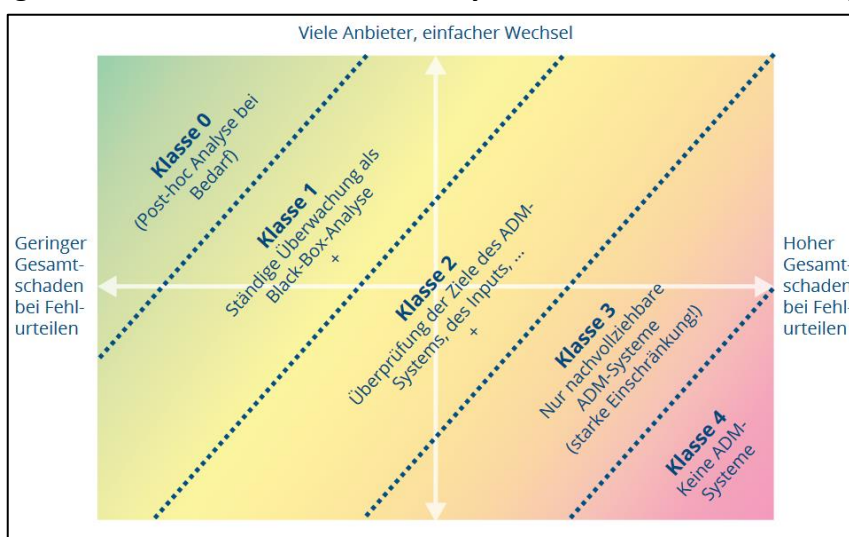
„Algorithmische Entscheidungssysteme beinhalten Algorithmen an zwei Stellen: Der erste Algorithmus lernt auf Basis der Daten ein statistisches Modell. Das statistische Modell ist dann die Grundlage für den (meist sehr einfachen) zweiten Algorithmus, der die eigentliche Entscheidung für eine neue Eingabe 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.“<sup>74</sup>

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 und wo welche Straftaten zu erwarten sind), Rückfalligkeitsvorhersagealgorithmen, Terroristenidentifikation.<sup>75</sup> Zweig nennt darüber hinaus mehrere Fehlerquellen bzw. Fehlerurteile 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 von Social Media 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“.<sup>76</sup>

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.

**Abbildung 6: Fünf Risikoklassen von ADM-Systemen mit lernenden Komponenten**



Quelle: Zweig (2019a) 13

Um die Fehlerquellen und die Risiken bei der Entwicklung und der Einbettung von algorithmischen Entscheidungssystemen zu minimieren, schlägt Katharina Zweig<sup>77</sup> einen Prozess in fünf Phasen vor.

1. Phase: Algorithmen werden designt 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 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.

#### **4.1.1. 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“.<sup>78</sup>

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 „Maschinlernen“. 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) Berater-Entscheidungen 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 Entscheidungen entsprechend adaptiert. Das Programm würde also aus den vergangene Entscheidungen lernen. Erste Ansätze zur Anwendung von schwacher künstlicher Intelligenz wurden in Kapitel 2.2. 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 Interessenslage der beteiligten Akteure wird das wohl auch noch lange so bleiben. Eigentlich schade.

#### **4.2. Ethische Fragestellungen und Menschenrechte**

„The digital welfare state is either already a reality or emerging in many countries across the globe. In these states, systems of social protection and assistance are increasingly driven by digital data and technologies that are used to automate, predict, identify, surveil, detect, target and punish. In the present report, the irresistible attractions for Governments to move in this direction are acknowledged, but the grave risk of stumbling, zombie-like, into a digital welfare dystopia is highlighted. It is argued that big technology companies (frequently referred to as “big tech”) operate in an almost human rights-free zone, and that this is especially problematic when the private sector is taking a leading role in designing, constructing and even operating significant parts of the digital welfare state. It is recommended in the report that, instead of obsessing about fraud, cost savings, sanctions, and market-driven definitions of efficiency, the starting point should be on how welfare budgets could be transformed through technology to ensure a higher standard of living for the vulnerable and disadvantaged.“<sup>79</sup>

Wie können digitale Technologien genutzt werden, um die Lebensverhältnisse von Arbeitslosen, Armen, Behinderten und anderen Hilfs- und Schutzbedürftigen zu verbessern – das sollte also lt. diesem UN-Bericht die eigentliche Fragestellung für den digitalen Wohlfahrtsstaat sein! Der Staat, also in unserem Fall die Institutionen des Sozialstaates, der Arbeitsmarktpolitik im Besonderen, sollten zuallererst rechenschaftspflichtig gegenüber den EmpfängerInnen von Transferleistungen wie dem Arbeitslosengeld und der Notstandshilfe sein und nicht umgekehrt (die BürgerInnen gegenüber dem Staat). In dem UN-Bericht werden eine Vielzahl an Beispielen für den digitalen Wohlfahrtsstaat, verteilt über den gesamten Globus, angeführt: Von der Identitätsprüfung (Indien, Kenia), der Überprüfung der Anspruchsvoraussetzungen für Sozialtransfers, der Berechnung der Auszahlungen, der Betrugsverhinderung, der Risikoeinstufung bis hin zur Kommunikation zwischen den Institutionen des Wohlfahrtsstaates und den TransferbezieherInnen. Laut Philip Alston sind durch den digitalen Wohlfahrtsstaat die verschiedenen Dimensionen der Rechte auf sozialen Schutz in Gefahr:

„While social protection in general should be designed to protect those rights, the dignity dimension is at particular risk in the context of the digital welfare state. The potential risks arise in various contexts.

First, the process for determining eligibility may easily be transformed into an electronic question-and-answer process that almost inevitably puts already vulnerable individuals at even greater disadvantage.

Second, the way in which determinations are framed and communicated may be dehumanized and allow no room for meaningful questioning or clarification.

Third, the digital welfare state often seems to involve various forms of rigidity and the robotic application of rules. As a result, extenuating circumstances, such as being late for an appointment because of urgent caring obligations or being unable to understand a written communication because of a disability or a personal crisis, are often not taken into account in a predominantly digital context.

Fourth, digital systems are often not designed to respond rapidly either to serious emergencies or to daily challenges, such as those that may be experienced by an older person whose entitlement has suddenly and inexplicably been electronically reduced or cancelled or by a single parent unable to take a child to a local day care because the digital identification card will not function.

Fifth, the ways in which services are provided can easily have degrading connotations, such as unnecessarily exposure to a broader audience the fact that a person is reliant on benefits, or requiring extended waiting periods or the navigation of lengthy queues.

Sixth, the introduction of various new technologies that eliminate the human provider can enhance efficiency and provide other advantages but might not necessarily be satisfactory for individuals who are in situations of particular vulnerability. New technologies often operate on the law of averages, in the interests of majorities and on the basis of predicted outcomes or likelihoods.

Seventh, digital services risk eliminating, almost entirely, much of the human interaction and compassion that are likely to be indispensable components in providing at least some welfare recipients with the care and assistance they need. The assumption that there is always a technological fix for any problem is highly likely to be misplaced in various aspects of a humane and effective system of social protection.<sup>80</sup>

In Bezug auf den AMS-Chancen-Modell sind diese Überlegungen in mehrfacher Hinsicht von Relevanz, insbesondere hinsichtlich des „law of average“, dem sich die Wirklichkeit häufig durch unerwartete Abweichungen zu entziehen sucht und der Annahme, es gäbe für jedes sozialpolitische Problem ein „technological fix“. Wenn die Maschine, der Computer einmal angeworfen wurden, folgt das Problem der De-Humanisierung auf dem Fuße!

Dabei muss betont werden, dass der Menschenrechtsansatz über die Forderung nach Transparenz und Rechenschaftspflicht weit hinausgeht bzw. viel tiefergehend ansetzt. Die Diskussionen dazu haben allerdings erst begonnen. In einem Bericht des Europarates aus dem Jahr 2018 werden die Auswirkungen von „automated data processing techniques“ auf verschiedenen Aspekte der Menschenrechte (Recht auf Privatsphäre und Datenschutz, wirksame Rechtsmittel, Diskriminierungsverbot, etc.) erörtert und regulatorische Implikationen diskutiert.<sup>81</sup> Auch das Gutachten der deutschen Datenethikkommission formuliert „Allgemeine ethische und rechtliche Grundsätze und Prinzipien“ im Hinblick auf Daten und Algorithmische Systeme.<sup>82</sup> Darüber hinaus geht ein Bericht zu Polen („Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making“)<sup>83</sup> sehr detailliert einigen Grundrechtsfragen zum Profiling, das 2014 eingeführt wurde, nach. Etwa dem Problem der mangelnden Transparenz – es kann von außen nicht genau nachvollzogen werden, wie der „Algorithmus“ (es handelt sich offensichtlich genauer genommen um eine logistische Regression) genau von den 24 Fragen des Fragebogens, die die Distanz zum Arbeitsmarkt und die Bereitschaft zur Wiedereingliederung messen sollen, zu dem Ergebnis der Einteilung in drei Gruppen kommt bzw. wie genau die Zuteilung zu den Programmen erfolgt. Dem

Problem, dass eine Korrektur des Profils, auch wenn sich die Lebensumstände (etwa Betreuungspflichten) geändert haben, schwer möglich ist. Auch der Gefahr der Stigmatisierung bzw. Diskriminierung v.a. hinsichtlich der Gruppe III, mit den schlechtesten Arbeitsmarktchancen, wird nachgegangen. Und schließlich wird das Problem thematisiert, dass einmal falsch dokumentierte Antworten im Fragebogen nur sehr schwer wieder korrigiert werden können.

Was bedeuten diese Überlegungen zu den grundrechtsrelevanten Aspekten von Algorithmen, algorithmischen Entscheidungssystemen und Künstlicher Intelligenz für das AMS-Chancen-Modell? Zunächst einmal wäre es wichtig, dass eine unabhängige Stelle – angesiedelt etwa beim Parlament oder bei einem wissenschaftlichen ExpertInnengremium, das nicht abhängig ist von Aufträgen des AMS oder des BMAFJ - die Kontrolle über die Entwicklung und Implementierung von AMAS ausübt, und die Transparenz und die Einhaltung aller Grundrechte der betroffenen Arbeitslosen sicherstellt! Es wäre wichtig, dass es zu einer umfassenden sozialwissenschaftlichen Begleitung kommt, die von den technischen Aspekten des Profiling ausgehend auch alle anderen Faktoren, die den Kontext der Entwicklung und der Implementierung betreffen, in den kritischen Blick nimmt. Einer Frage, die sich dieses ExpertInnengremium u.a. annehmen sollte, wäre jene nach den möglicherweise diskriminierenden Strukturen, die in den von Synthesis verwendeten Datensätze eingeschrieben sind: Sind die HV-Daten und die AMS-Registerdaten völlig frei von diskriminierenden Aspekten, in dem Sinn, dass alleine die Art der Aufbereitung, der Fragestellungen etwa zum bisherigen Erwerbsverlauf von Frauen, MigrantInnen, behinderten Menschen, etc. für die Ermittlung der Profiling-Ergebnisse im Effekt(!) diskriminierend wirken. Bisher hat noch niemand diese Fragen an die Daten gerichtet! In diesem Zusammenhang sollte etwa auch das Problem behandelt werden, was es heißt, dass Synthesis ja „nur“ Korrelationen ermittelt, diese allerdings kausal interpretiert werden; etwa: wenn eine arbeitslose Person in einem Arbeitsmarktbezirk mit schlechten Arbeitsmarktchancen wohnt, dann sind die Wiederbeschäftigungschancen dieser Person schlechter zu bewerten. Hinzu kommen weitere Fragen: Etwa jene, nach dem Vorgehen, wenn eine Person durch das Profiling falsch zugeordnet wurde (immerhin möglicherweise 20-40% aller Arbeitslosen, siehe Kapitel 2.4.1.). Wie kann dieser Fehler ermittelt werden, wie können sich die Betroffenen in diesen Fällen zur Wehr setzen. Wie wird überhaupt mit dem Problem der statistischen Diskriminierung umgegangen?

Darüber hinaus sollten die BeraterInnen, also die Personen, die AMAS dann endgültig umsetzen müssen, aktiv in die Entwicklung und v.a. die Umsetzung eingebunden werden. Schließlich muss den Arbeitslosen, als die eigentlich Betroffenen, ein umfassendes Informationsrecht eingeräumt werden: Wie sind die Ergebnisse zu den Wiederbeschäftigungschancen des/der betroffenen Arbeitslosen konkret zustande gekommen, welche Faktoren wirkten positiv, welche negativ? Das wäre dann ein erster Schritt, um die grundsätzlichen Machtasymmetrien zwischen AMS-BeraterInnen und den Arbeitslosen auszugleichen. Das müsste allerdings verbunden sein mit einem allgemeinen Kulturwechsel im AMS: Dem paternalistischen Verständnis des AMS in Bezug auf ihre KundInnen, nämlich den Arbeitslosen, sollte ein aufgeklärter und (menschen-)rechtsstaatlicher Zugang entgegengesetzt werden! Das hätten dann allerdings weitreichendere Konsequenzen, die über das AMAS hinausreichen würden. Etwa, wenn den Arbeitslosen das Recht auf eine wirksame Maßnahme eingeräumt werden würde, also keine „Sinnloskurse“ mehr. Es sollte die Mitsprache bei

der Maßnahmen-Entscheidung bzw. Vorschlagsrecht der Arbeitslosen wie bei den Kurskosten eingeräumt werden –und: die Ablehnung durch BeraterIn müsste begründungspflichtig sein!

## 5. Schlussfolgerungen

Aus den vorangegangenen Ausführungen zum AMS-Profilng (AMAS) sollten 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 für besondere Problemgruppen verbessert werden? – nichts beitragen kann.
- Die internationalen Erfahrungen mit Profiling seit den frühen 1990er Jahren zeigen nämlich, 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 bis nichts 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 an Transparenz haben mehrfach ForscherInnen kritisiert, leider wurden daraus von Seiten des AMS und der Aufsichtsbehörde, dem BMAFJ, noch keine Konsequenzen daraus gezogen.
- Das AMS hat 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 Vermittlungshemmungen mehr Beratung und Unterstützung (BBEN) von Trägern außerhalb des AMS bekommen und teurere Kurse nicht mehr bekommen sollen. Dafür hätte es allerdings das Profiling nicht gebraucht.
- Richtig ist jedoch, 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; es handelt sich nämlich um eine einfache logistische Regression, gepaart mit viel Dezisionismus an allen Ecken und Enden. Aber aus einem grundsätzlicheren Blickwinkel wären die Voraussetzungen für den Einsatz von Künstlicher Intelligenz (Stichwort: Big Data) in der Arbeitsmarktpolitik eigentlich gut.
- Die Anwendung von künstlicher Intelligenz in der Arbeitsmarktpolitik wäre allerdings nur dann zu befürworten, wenn anders als bisher vorgegangen werden würde: Dazu wäre es unbedingt notwendig, die Entwicklung und Implementierung von Künstlicher Intelligenz mit einer umfas-

senden sozialwissenschaftlichen Begleitung und einer Technologiefolgenabschätzung zu verbinden, 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.

## 6. 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*.  
<https://www.frontiersin.org/articles/10.3389/fdata.2020.00005/full>
- Alston, Ph., Report oft he 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](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](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](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, 24.04.2019).  
<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](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](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. Ujohn 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](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-datsaetze?detail-view=yes&publikation\\_id=50691](https://www.wifo.ac.at/pubma-datsaetze?detail-view=yes&publikation_id=50691)
- Eppel, R., Leoni, Th., Mahringer, H., Hausegger, T., Reidl, Ch., Weber, F., Einsatz und Wirkung aktiver arbeitsmarktpolitischer 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](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.cms-wire.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. 09. 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 18<sup>th</sup> Conference in Graz 2019 (6<sup>th</sup>-7<sup>th</sup> 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](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, 01.02.2019). [http://www.forschungsnetzwerk.at/downloadpub/PPC Endbericht abif.pdf](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](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/pri3/wifo/main.iart?content-id=1454619331110&publikation\\_id=28498&detail-view=yes](https://www.wifo.ac.at/jart/pri3/wifo/main.iart?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](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 (Brussels 2001).  
[www.pls-ramboll.com/homepage/uk/publications/](http://www.pls-ramboll.com/homepage/uk/publications/)
- Spiekermann, S., Warum das AMS keine KI auf österreichische Bürger loslassen sollten, in: der Standard (23. 09. 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 (27.09.2019).  
<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](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](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](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](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 (Berlin, 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. <https://www.oeffentliche-it.de/documents/10181/76866/8+Zweig+%26+Krafft+-+Fairness+und+Qualität+algorithmischer+Entscheidungen.pdf>

## 7. Anmerkungen

<sup>1</sup> 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 aufzubauchen oder in einer bestimmten Zeit wieder einen Job 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; findet etwa in Estland und Deutschland Anwendung; „Rules-based profiling“, hier werden einfache Regeln angewendet, etwa, dass Personen mit einem bestimmten Alter oder einer bestimmten Arbeitslosigkeitsdauer in ein Programm kommen können; Beispiele dafür sind Norwegen und Polen; schließlich „data-assisted profiling“, das ist eine Mischform aus Beraterentscheidungen und statistischer Unterstützung; Loxha und Morgandi (2014) 13.

<sup>2</sup> Ein Algorithmus ist eine Abfolge von Befehlen, eine Handlungsanleitung, eine Zuordnungsregel ähnlich einem Kochrezept: „Man nehme ...“. Siehe weiterführend im Detail unter 4.1.

<sup>3</sup> Beispielsweise Berger et. al. (2001).

<sup>4</sup> Szigetvari (2019), Kopf (2019a, c).

<sup>5</sup> Für einen frühen Versuch zur Bestimmung der individuellen Dauer der Arbeitslosigkeit im österreichischen AMS siehe Dornmayr (1996).

<sup>6</sup> OECD (1998). In der Einleitung zu dieser Publikation heißt es im ersten Satz dann noch deutlicher: “This Report examines the early identification of jobseekers at risk of becoming long-term unemployed (known as ‘profiling’)”. OECD (1998) 3. 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.

<sup>7</sup> OECD (1998) 9.

<sup>8</sup> Grundsätzlich lassen sich folgende Allokationsmechanismen unterscheiden: 1. Entscheidung des AMS-Beraters; 2. Deterministisch - etwa alle Jugendliche oder LZA erhalten eine Förderung; 3. „*First-come-first-serve*“-Regel; 4. Zufällige Zuteilung; und 5. Statistisches Profiling. Die zwei wesentlichen Zielsetzungen sind Effizienz der Maßnahmenteilnahme und Gerechtigkeitsüberlegungen (die am meisten Benachteiligten sollen besonders gefördert werden). Nur in Ausnahmefällen sind diese beiden Ziele ident!

<sup>9</sup> 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.

<sup>10</sup> Diese Wiedereingliederungsmaßnahmen beinhalten v.a. weniger intensive Beratungs-, Betreuungs- und Vermittlungsaktivitäten: „Services may include counseling, job search workshops, testing, job referral and placement, or if needed, referral to more intensive services, such as training“. Nur 13,8% der BezieherInnen von Arbeitslosengeld wurden zu teureren und intensiveren Programmen wie Qualifizierung und Training vermittelt. Black et.al. (2003) 1315, 1317.

<sup>11</sup> Für Großbritannien liegen aus dieser Zeit ähnliche Befunde vor: „In making a general point on referrals, the United Kingdom noted that the indicators which might relate to the risk of long-term unemployment, e.g. age, sex, etc. might not bear any relationship to the type of ALMP that is necessary. For example, although various variables may be associated with an at-risk group, they may be proxying different impediments to getting work. Some may relate to motivation, others to economic cycle, others to household characteristics, etc. To actually determine the appropriate ALMP requires more in-depth analysis than simply referring all jobseekers at risk to similar programmes.“ OECD (1998) 21.

<sup>12</sup> 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.

<sup>13</sup> 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 Prozent-Anteils an Leistungsbezug in einer Zeitperiode von 30 Monaten nach Beobachtungsperiode

<sup>14</sup> 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).

<sup>15</sup> Desiere (2019) 6f.

<sup>16</sup> Langenbacher (2019).

<sup>17</sup> Desiere et. al. (2019) 13.

<sup>18</sup> "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.

<sup>19</sup> Desiere (2019) 24.

<sup>20</sup> Gibbons (2018) 8. Siehe auch Loxha und Morgandi (2014) 20.

<sup>21</sup> OECD (2018) 3.

<sup>22</sup> Desiere et.al. (2019) 15 und Arni und Schiprowski (2015) 22ff.

<sup>23</sup> Klewais (2018), Desiere (2019).

<sup>24</sup> Wijnhoven und Havinga (2014) 744.

<sup>25</sup> Dornmayr 1996, Synthesis 2007.

<sup>26</sup> Synthesis 2016 a, b, 2017, 2018, 2019.

<sup>27</sup> Synthesis 2018.

<sup>28</sup> 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 „PAMAS“ bzw. „AMAS“ bezeichnet, die Abkürzungen für „(Personalisiertes) Arbeitsmarktchancen-Assistenz System“.

<sup>29</sup> Siehe nachfolgenden Screenshot aus Synthesis 2018, 11.

<sup>30</sup> 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.

<sup>31</sup> 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.

<sup>32</sup> Junger Mann, maximal Pflichtschule, Österreicher, keine Betreuungspflichten, nicht gesundheitlich beeinträchtigt, wohnhaft etwa in einem Arbeitsmarktbezirk wie Bregenz oder Amstetten, vor Arbeitslosigkeit beschäftigt im Dienstleistungssektor.

<sup>33</sup> Vergleich Synthesis (2019a) 16 mit Synthesis (2018) 15.

<sup>34</sup> Kopf (2019c).

<sup>35</sup> Cech et.al. (2019); und aktuell auch Allhutter et.al. (2020).

<sup>36</sup> Synthesis 2018.

<sup>37</sup> Kopf (2019c).

<sup>38</sup> Kopf (2018b).

<sup>39</sup> Diese Diskussionen mit Synthesis werden seit langer Zeit geführt und sind in der Community der ArbeitsmarktforscherInnen hinreichend bekannt. Unverständlich in diesem Zusammenhang ist der Umstand, dass dieses Unternehmen eine europaweite Ausschreibung gewinnen konnte, bei der so namhafte Forschungsinstitute wie etwa das WIFO gerade bei der „wissenschaftlichen Eignung“ schlechter als Synthesis abgeschnitten haben. Das Ausschreibungsverfahren wurde jedoch nicht veröffentlicht, denn in diesem Fall hätte sich das AMS (der Vorstand) rechtfertigen müssen, wie es zu dieser Vergabeentscheidung, die immerhin € 237.142,80 ausmachte, kommen konnte. Siehe Wimmer (2018). Auf Grund einer parlamentarischen Anfrage mussten die Zahlen für die Entwicklungskosten offensichtlich kräftig nach oben korrigiert werden, nämlich auf € 1,8 Mio.; siehe Szgetvari (2020).

<sup>40</sup> 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.

<sup>41</sup> 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."

<sup>42</sup> Synthesis (2019b) 10.

<sup>43</sup> 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).

<sup>44</sup> Synthesis (2018) 15. Im internationalen Vergleich ist das eine sehr hohe Trefferquote (siehe die vorangegangene Tabelle 2), wobei Zweifel berechtigt sind, warum gerade Österreich so viel besser abschneidet als viele andere Länder.“; Buchinger (2019), Frage 13.

<sup>45</sup> Desiere (2019) 24.

<sup>46</sup> Sendhil Mullainathan, The New York Times, Dec. 6, 2019.

<sup>47</sup> Eberts et.al. (2002).

<sup>48</sup> Kopf (2018).

<sup>49</sup> Kopf (2019a).

<sup>50</sup> Es ist die Rede von 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.

<sup>51</sup> In diesem Falle werden Beschäftigungsaufnahmen nach Maßnahmenende mit kausalen Wirkungen des Programmes verwechselt; siehe weiterführend unter 3.2.1.

<sup>52</sup> AMS-Verwaltungsrat (2019), Szigetvari (2019).

<sup>53</sup> AMS-Verwaltungsrat (2019) 3.

<sup>54</sup> Siehe Liebeswar et.al. (2019).

<sup>55</sup> Buchinger (2019) Frage 7.

<sup>56</sup> 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.

<sup>57</sup> Weinberger (2019).

<sup>58</sup> Das ist zwar keine Kontrollgruppenanalyse, aber es werden zum selben Zeitpunkt dieselben Problemgruppen für zwei Maßnahmen analysiert. Das kommt einem Matching recht nahe.

<sup>59</sup> Weber et.al. (2019).

<sup>60</sup> 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.

<sup>61</sup> Bei den Befragten waren zwei Drittel Männer, zwei Drittel hatten maximal einen Pflichtschulabschluss, die Mehrheit war über 50 Jahre alt, 80% hatten gesundheitliche Einschränkungen oder eine begünstigte Behinderung und alle waren mindestens 2 Jahre arbeitslos; Weber et.al. (2019) 9.

<sup>62</sup> Böheim et.al. (2020). In dieser Untersuchung wurden die Zielgruppen mit multiplen Vermittlungshemmnissen (siehe vorangegangene Fußnote) in Pilot-Regionen mit BBEN und ohne BBEN 2018-2019 verglichen in einem Zeitraum von 1,5 Jahren.

<sup>63</sup> Hier sollen nur Evaluierungen behandelt werden, die mit Kontrollgruppen (Matching) arbeiten. Für einen Überblick siehe etwa Eco-Austria (2016) und Schweighofer (2013).

<sup>64</sup> Lutz et.al. (2005), Lechner et.al. (2007).

<sup>65</sup> Eppel et.al. (2011).

<sup>66</sup> Lechner et.al. (2007) 56.

<sup>67</sup> Eppel et.al. (2014) 58.

<sup>68</sup> Eppel et.al. (2017).

<sup>69</sup> Behncke et.al. (2010) 6.

<sup>70</sup> Behncke et.al. (2010) 18.

<sup>71</sup> 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.

<sup>72</sup> Zweig (2019a) 3.

<sup>73</sup> Zweig (2019b) 60f.

<sup>74</sup> Zweig (2019a) 4.

<sup>75</sup> Lischka und Klingel (2017), hier sind weitere internationale Fallbeispiele für algorithmische Entscheidungssysteme zu finden.

<sup>76</sup> Zweig (2019a) 7.

<sup>77</sup> Zweig (2018).

<sup>78</sup> Niranjan Krishnan, Tiger Analytics, zitiert nach Ismail (2018).

<sup>79</sup> Alston (2019) 1.

<sup>80</sup> Alston (2019) 17.

<sup>81</sup> Council of Europe (2018).

<sup>82</sup> Datenethikkommission (2019).

<sup>83</sup> Niklas (2015).