

FramTEST: Framtidens testmetodik - Förstudie - Publik rapport



Författare: Anders Hjalmarsson Jordanius (Redaktör)
Datum: 2019-01-20
Projekt inom Elektronik Mjukvara och Kommunikation

FFI Fordonsstrategisk
Forskning och
Innovation

VINNOVA

Energimyndigheten

TRAFIKVERKET

FKG

VOLVO

SCANIA

VOLVO

SCANIA

VOLVO

Innehållsförteckning

1 Sammanfattning	3
2 Executive summary in English	3
3 Bakgrund	3
4 Syfte, forskningsfrågor och metod	4
5 Resultat och måluppfyllelse	6
5.1 Fas 1: granskning av state-of-art.....	6
5.2 Fas 2: explorativ intervjustudie	10
5.3 Fas 3: analys och syntes	12
6 Spridning och publicering	13
6.1 Kunskaps- och resultatspridning	13
6.2 Publikationer	14
7 Slutsatser och fortsatt forskning	14
8 Deltagande parter och kontaktpersoner	14

Kort om FFI

FFI är ett samarbete mellan staten och fordonsindustrin om att gemensamt finansiera forsknings- och innovationsaktiviteter med fokus på områdena Klimat & Miljö samt Trafiksäkerhet. Satsningen innebär verksamhet för ca 1 miljard kr per år varav de offentliga medlen utgör drygt 400 Mkr.

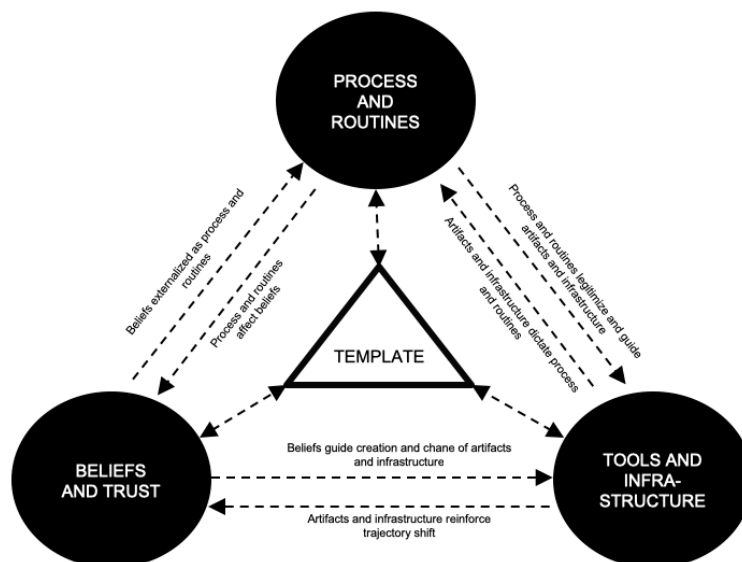
För närvarande finns fem delprogram; Energi & Miljö, Trafiksäkerhet och automatiserade fordon, Elektronik, mjukvara och kommunikation, Hållbar produktion och Effektiva och uppkopplade transportsystem. Läs mer på www.vinnova.se/ffi.

1 Sammanfattning

Förstudien FramTEST har inriktats mot att identifiera mekanismer som möjliggör att ny testmetodik baserat på AI, machine learning och automatisering kan implementeras för validering och verifiering av ny mjukvara i fordonssystem. Studien har genomförts under 2018 och innefattat 1) en genomgång av 23 projekt inom FFI programmet EMK som klassats som verifiering och valideringsprojekt 2) en litteraturanlys av 73 artiklar inom området 3) en explorativ intervjuserie med 12 respondenter inom fordonsindustrin och närliggande branscher för att fånga hur de som experter inom testning ser hur ny testmetodik ska implementeras. Projektet har tagit fram ett förslag på forskningsmodell för att förklara och förstå hur olika faktorer samspelar när ny datadriven testmetodik för mjukvaruutveckling implementeras i fordonsindustrin. Förstudien har därmed skapat förutsättningar för fortsatt forskning att ta fram kunskap och strategi för hur datadrivna testmetoder införs och anammas av etablerad industri.

2 Executive summary in English

FramTEST as a pre-study has been aimed to identify mechanisms that enable organizations within the automotive industry to embed novel techniques for software validation and verification based on AI, machine learning and increased automatization. Such techniques are viewed as key to enable the automotive industry to develop secure and safe vehicles that builds on the vision of autonomous transport. The pre-study was performed during 2018 and included three phases. In the first phase, state-of-art, a meta-review was done on 23 projects funded by FFI and classified as validation and verification. In addition, a literature review of 73 papers was performed to elicit potential factors that affect the transformation of the validation process within an incumbent organization towards data-driven verification. The state-of-art review generated a model with three socio-technical categories that interplay when building a data-driven validation culture, with accompanying suggestions of sub-factors. Through an explorative semi-structured interview phase the research model was partly-grounded in empirical data. This resulted in the addition of additional sub-factors, the identification of a potential core variable, as well as a base to develop propositions about the dependencies between the variables. The pre-study has created the basis for two scientific papers and a model that could guide future research how data-driven validation models can be implemented and embedded in the incumbent automotive industry.



3 Bakgrund

Fordonsindustrin är draglok för svensk export och står för en betydande del av den svenska industrins sysselsättning och förädlingsvärde. Tre trender med den gemensamma nämnaren

mjukvarubaserade funktioner stakar ut förutsättningarna för fordonsindustrins framtida utveckling: elektrifiering, automatisk funktionalitet och mobilitetstjänster. Men innovation av mjukvarubaserade funktioner leds och organiseras annorlunda än hur innovation traditionellt skett inom fordonsindustrin. Fordonsindustrins innovation bygger dels på hårt driven uppdelning av fordon i delsystem, dels på delat ansvar för innovation mellan fordonstillverkare och underleverantörer. I och med elektrifiering, autonom fordonsfunktionalitet och mobilitetstjänster ökar såväl mängden mjukvara som beroendena mellan fordonens olika delsystem och uppdelningen blir ohållbar. Detta innebär också att renodlade mjukvaruutvecklare, som tidigare varit sällsynta blir alltmer intressanta som underleverantörer till fordonsindustrin parallellt med att industrin fortsätter samarbetet med större leverantörer av helhetslösningar.

För att svensk fordonsindustri ska lyckas behålla sin framskjutna ställning krävs att fordonstillverkare och dagens underleverantörer stärker sin förmåga att effektivt testa mjukvara för framtidens fordon. Under innevarande årtionde har teknikutvecklingen kring maskininläring (ML) och artificiell intelligens (AI) accelererat. En fjärde industrirevolution står för dörren med ökad grad av automatisering och samverkande system där ML och AI är nyckelkomponenter. Utökad forskning krävs för att bättre förstå hur ML och AI kan bidra till att testning av mjukvara i framtidens fordon kan effektiviseras. Att styra om hur testning sker i fordonsindustrin kan innebära stora förändringar och utmanar förhärskande föreställningar hur företag organiserar sina utvecklings- och testprocesser. Nya former behövs för hur aktörer genom öppna, flexibla och automatiserade processer samverkar för att verifiera och validera mjukvaruinnovationer där tid till marknad är kritisk och samtidigt systemen är tillförlitliga och säkra (Juell-Skielse & Hjalmarsson-Jordanius 2017).

4 Syfte, forskningsfrågor och metod

Föreliggande förstudie - **FramTEST** - har förankrats i följande forskningsfråga:

Hur bör testmetodik utformas som möter framtidens krav på verifierade och validerade fordonssystem?

Syftet har varit att karaktärisera existerande kunskap om metoder och praktiker inom fordonstestning och kartlägga state-of-art forskning inom området. Av särskilt intresse har varit att förstå hur automatisering, AI och ML system kan förbättra testning av komponenter som i sin tur skapar autonoma och självlärande transport- och fordonssystem. En viktig del har därför varit att inledningsvis granska projekt som av FFI finansierats inom klassificeringen verifiering och validering för att härigenom förstå de fokus som tidigare realiseringsprojekt har haft i förhållande till AI, ML och automatisering. Med detta som grund har en större litteraturstudie utförts för med syftet att förstå kunskapsfronten för testning och validering inom fordonssektorn.

Med utgångspunkt i state-of-art granskningen har projektet även omfattat målet att explorativt fånga och förstå domänexperters syn på behov av framtida testtekniker och processer baserat på AI, ML och automatisering. Fokus under intervjuerna har i dialog med respondenterna även kommit att inkludera förutsättningar för att styra om nuvarande former för testning och således kommit att inkludera barriärer för att implementera ny metodik och bygga en kultur för testning där ansvaret för validering och verifiering i högre grad än tidigare fränkopplas mänsklig bedömning.

För att möjliggöra kompetenshöjning och kunskapsutbyte har detta förstudieprojekt utförts i samarbete mellan en industripart (Infotiv AB), med fokus på testning och validering, och ett industriforskningsinstitut (RISE Viktoria). Parterna har samarbetat i fyra arbetspaket (se figur 1). Inledningsvis har en granskning av FFI/EMK finansierade projekt utförts. Därefter har en undersökning av internationell forskning inom validering och verifiering genomförts (AP2). Resultatet av state-of-art medförde att fokus riktades om från hur testmetodik utformas till hur testmetodik för att möta framtida krav på verifiering och validerade fordonssystem implementeras.

En explorativ intervjustudie har därefter skett i AP3 med syfte att förstå domänexperters syn på framtida tekniker och processer och utmaningar som dessa medför på testpraktiken. De experter som har intervjuats representerar den svenska fordonsindustrin och industrier som redan påbörjat använda tekniker och processer för testning baserat på ML, AI och automatisering. Baserat på en syntes av resultaten från AP2 och AP3 har en mappning av behov och utmaningar utförts som

pekar på tre områden som behöver adresseras när framtida testmetodiker utvecklas och implementeras inom fordonsindustrin (AP4).



Figur 1. Arbetspaketsstruktur i förstudien **FramTEST**

För att genomföra AP2 tillämpades två olika metoder under projektets inledande tre månader (Mars-Juni 2018). Analysen av genomförda projekt inom FFI / EMK har väglett av det underlag som programkontoret har tillhandahållit projektet. Samtliga projekt i programmets projektkatalog har klassificerats. De projekt som hamnat under klassen verifiering och validering, totalt 23, har därefter analyserats och ytterligare klassificerats. Klassificeringen har utförts genom tre kriterier 1) relevans (i förhållande till **FramTEST** forskningsfrågor), 2) projektets fokus i förhållande till testning, 3) explicit koppling till AI, ML och automatisering av verifiering och validering. Genom denna gallring identifierades 13 projekt av särskilt intresse för **FramTEST**. De projekt som under förstudien pågick har kontaktats för att erhålla underlag och information medan de projekt som avslutats har studerats i form av dokumentation som har varit tillgängligt. Den därpå efterföljande litteraturstudien utgick dels från resultatet av analysen av de utvalda projekten, dels från den systematiska teknik för litteraturstudier som presenteras i Wolfwinkel et al. (2013). Av särskilt värde för projektet var den koppling som kunde göras mot det pågående projektet SMILE II. Inom ramen för detta EMK-projekt hade en litteratursammanställning utförts vilkas referenser tillhandahålls FramTEST. Totalt under litteraturstudien granskades 71 artiklar mellan perioden 2006 och 2018.

Granskningen av projekt inom EMK med fokus på verifiering och validering samt utförandet av litteraturstudien avslutades i juni 2018. I anslutning till uppstarten av AP3 genomfördes en kunskapsspridningsaktivitet där delresultat presenterades vid ett frukostseminarium arrangerat av Infotiv (se avsnitt 6.1). Med utgångspunkt i AP2 utformades under augusti en semi-strukturerad intervjuguide för att fånga hur domänexperter ser på testning i framtiden inom fordonsindustrin. Intervjuguiden skapades genom två workshops med projektdeltagare där även forskare från SMILE bjöds in att bidra. Guiden kom totalt att omfatta sju intervjuteman och planerades att ta ungefär 60 minuter att genomföra tillsammans med respondenten. I möjligaste mån planerades att två personer från projektet, en från RISE och en Infotiv, deltog i varje intervju för att frågorna skulle kunna belysas från både ett forsknings- och ett industriperspektiv. Genom detta skapades förutsättningar för triangulering vid datainsamlingstillfällena. En lista med 16 respondenter identifierades och kontaktades för intervjuer. Av de tänkta 16 respondenterna hade slutligen 12 respondenter möjlighet att ställa upp på intervjuer i **FramTEST**.

Intervjuerna genomfördes under perioden september till november. Löpande transkriberades inspelningarna av personal från Infotiv och ställdes samman till 12 intervjumemos. Analysen av intervjuerna följde en två-stepsprocedur (Clarke & Braun 2013). Inledningsvis analyserades samtliga teman genom enskild analys av 12 intervjuerna. Därefter jämfördes insikterna från respektive intervju med materialet som helhet. Ur materialet identifierades även aspekter rörande implementering av testmetodik som inte genom temana hade förutsetts i planeringen av

intervjuerna. Med andra ord tillämpades både tilldelande och friläggande tolkning av materialet som samlats in. För att underlätta kodning och jämförelse mellan intervjuerna har projektet tillämpat atlas.ti. för en inledande kvalitativ analys av datamaterialet (Woolf & Silver 2018). Detta arbete kommer att fortsätta, bland annat genom att projektet tillhandahåller data till det pågående SMILE II projektet.

Arbetet i AP3 att genomföra intervjuerna, transkribera materialet och göra en inledande analys av intervjuerna utfördes under september-november 2018. Under december har resultatet från AP2 och AP3 jämförts och en övergripande modell har utformats med exempel på behov och krav för implementering av framtida testmetodik inom fordonsindustrin. Under december färdigställdes även slutrapporten för förstudien.

5 Resultat och måluppfyllelse

Projektet har delat in resultaten i de tre huvudfaser som utgjort arbetet. I detta avsnitt introducerar vi övergripande studiens resultat. Resultaten utgör grund för två vetenskapliga publikationer som under 2019 kommer sändas in för bedömning i två vetenskapliga outlets (se avsnitt 6.2).

5.1 Fas 1: granskning av state-of-art

Fas 1 omfattade två delmoment. Dels genomfördes en kartläggning av avslutade och pågående FFI-projekt inom programmet EMK som är klassade som verifiering och validering.

Kartläggning FFI-projekt inom EMK verifiering och validering

23 projekt utav samtliga projekt inom EMK klassificerades som verifierings- och valideringsprojekt. Av dem var sex projekt pågående och resterande projekt avslutade. Projekten som granskades var beviljade under perioden 2011-2017 med huvuddelen av projekten beviljade från 2015 och framåt. Ett syfte med kartläggningen var att få kontakt med aktörer inom projekten som å ena sidan kunde tillföra projektet underlag för state-of-art-analysen, å andra sidan identifiera projekt som kan berikas av förstudiens resultat. Ett andra syfte med kartläggningen var att fånga och förstå de projekt som särskilt adresserar ML, AI och automatisering vid testning, vilket utgjort **FramTEST** fokus.

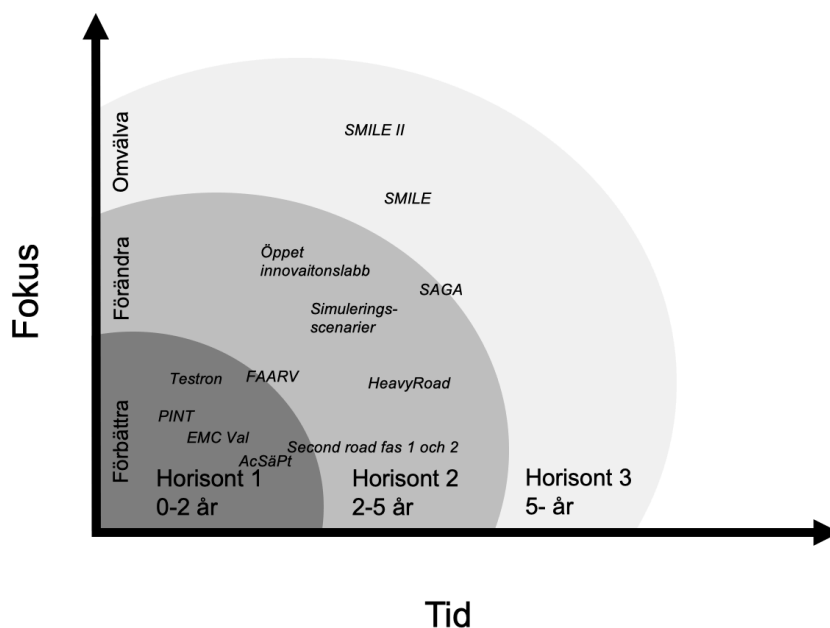
Diarium	Namn	Kartläggning av FFI-projekt inom verifiering och validering i relation till FramTEST syfte		
		Grad av relevans för FramTEST	Projektets fokus i relation till testning	Grad av koppling till ML och AI driven testning
2016-05497	Öppet innovationslabb (**)	Hög	Testmiljö inkl simulering	Låg
2016-05495	Simuleringsscenarier (**)	Hög	Simulering	Låg
2016-05492	Virtuell Motor Kalibrering (VirCal)	Låg	Virtuell testmiljö	Låg
2016-05399	BRAVE -- Laddtryckreglering av motorer med avancerade ventildrivtåg	Låg	Annat	Låg
2016-05380	FROST: FörbrukningsReduktion - Optimal Strategi och verkTygslåda	Låg	Annat	Låg
2016-04255	SMILE - Säkerhetsanalys och verifiering/validering av system baserade på maskininläring (***)	Hög	ML / AI	Hög
2017-03066	SMILE II - Säkerhetsanalys och verifiering/validering av system baserade på maskininläring (***)	Hög	ML / AI	Hög
2016-04238	SEER - Sömlös, Effektiv och bEhaglig inteRaktion (SEER)	Låg	Annat	N/A
2016-02560	FAARV - Feasibility of testing relevant ADAS and AD scenarios in an Augmented Reality Vehicle (*)	Medel	Virtuell testning	Låg
2016-02496	HiFi Visual Target	Låg	Metodutv.	Låg
2016-02487	Modellering av uppvärmning/nedkylning av hybridkomponenter p.g.a. omgivande luftflöden.	Låg	Modell	Låg
2015-06892	EMC Validering av multi-sensorsystem (*)	Medel	Metodutv.	Låg
2015-04893	Modellbaserad Testning av Mekatroniska System (TESTRON) (*)	Medel	Modell	Låg
2015-04852	HiFi Radar Target	Låg	Simulering	Låg
2015-04816	SAGA - Situationsbaserad Integrationstestning av Fordonssystem med hjälp av Guarded Assertions (***)	Hög	Automatisering	Medel
2014-03947	HeavyRoad (**)	Medel	Simulering	Medel
2011-04438	ISO 26262 Verifiering av programvara i säkerhetskritiska EE-system (*)	Medel	Simulering	Låg
2011-00468	Second Road fas 1 - En gemensam simulatormiljö för aktiv säkerhet och HMI inom Volvo Personvagnar (*)	Medel	Simulering	Låg
2009-01052	Reducering av ledtid för virtuell provning inom fordonsindustrin	N/A	N/A	N/A
2013-04727	Second Road fas 2 (*)	Medel	Simulering	Låg

2013-01296	Synligare utveckling av inbyggda fordonssystem - visualiserad kravhantering och samverkan	Låg	Annat	Låg
2013-01283	Parallellisering av integrationstest (PINT) (*)	Medel	Metodutv.	Låg
2012-00943	AcSäPt - Acceptanstest av säkerhetskritisk plattformsprogramvara (*)	Hög	Modell	Låg

Tabell 1. Värdering av EMK projekt inom domänen verifiering och validering

För att effektivisera kartläggningen och samtidigt säkra en konstruktiv output från arbetet till senare steg i förstudien definierades tre kriterier för att karaktärisera inriktningen på de olika projekten och bedöma relevansen med projektet för **FramTEST**. Av projekten har/hade sex projekt ett uttalat simuleringsperspektiv. Av de sex projekten fanns tydlig koppling till AI, ML och automatisering i ett av projekten (HeavyRoad). Ytterligare fyra projekt, SecondRoad fas 1 och, ISO 26262 Verifiering, Simulerings-scenarier värderades som relevanta och granskades närmare. Närliggande fokus till simulering är virtuell testningsmiljö vilket två projekt klassades som. Båda projekten erhöll låg explicit koppling till AI, ML och automatisering, emellertid bedömdes FAARV genom projektets fokus som relevant för granskning inom ramen för **FramTEST**.

Metod- och modellutveckling har adresserats i sex projekt finansierade av programmet. Inget av projekten har explicit koppling till AI, ML och automatisering. För **FramTEST** identifierades att AcSäPt, PINT, EMC Validering samt TESTRON genom deras fokus på att ta fram metoder för avancerad testning inom fordonsfältet som relevanta för närmare granskning i förstudien. De projekt som genom karaktäriseringen fick högst relevans för **FramTEST** var å ena sidan SAGA vilket hade ett fokus på automatisering av testprocessen med viss explicit relatering till AI och ML samt de två projekten SMILE och SMILE II som har hög grad av explicit koppling till AI och ML vid testning inom fordonsindustrin. SMILE II är pågående vilket föranledde **FramTEST** att kontakta med forskningsteamet för att säkra en koppling mellan **FramTEST** som förstudie och SMILE II som realiseringsprojekt.



Figur 2. Positionering av projekt inom EMK mot ambitionen att utveckla testning

FramTEST är en förstudie för att studera de mekanismer som krävs för att implementera framtida testmetodiker inom fordonsindustrin. I arbetet med att analysera de projekt som valts ut har projektbeskrivningar, slutrapporter, resultat (t.ex. artiklar) samt i de fall projektedeltagare haft möjlighet att ställa upp intervjuas. Centralt för **FramTEST** har varit att matcha förstudiens ambition att studera vad som krävs för att omvärva testverksamhet med de genomförda projektens ambition. Ambitionen har delats in i tre utvecklingshorisonter: förbättra nuvarande verksamhet, förändra nuvarande verksamhet samt omvärva hur testning genomförs inom verksamheten. Figur 2 visar den positionering som genomfördes av de projekt inom EMK som valdes ut. Förstudien kunde konstatera att majoriteten av projekten har haft ett fokus att genom sina resultat påverka den pågående testverksamheten. Endast tre projekt har explicit inriktats mot att studera hur testning

kan genomföras genom "disruptiv" omvälvning. SMILE och pågående SMILE II definierades som viktiga källor men också mottagare av resultat från **FramTEST** som förstudie.

Litteraturstudie

Genom att studera resultaten från SAGA och SMILE skapades förutsättningar för att rikta in den litteraturstudie som utgjorde **FramTEST** andra steg i fasen State-of-Art. Syftet med SMILE var att genom en inledande litteraturstudie skapa förutsättningar för en framgångsrik workshopserie där behov från industrin kopplad ML baserad verifiering diskuterades. **FramTEST** tog i sitt urval av litteratur, för att utforska mekanismer för att framgångsrikt implementera testmetodik baserat AI och ML, utgångspunkt i det litteraturval som skett inom SMILE. Detta för att skapa förutsättningar för att forma en gemensam plattform för senare utbyte mellan **FramTEST** och SMILE II. Som tillägg adderades litteratur genom liknande urvalskriterier vilka inte var publicerade under SMILE. Det urval som kunde genomföras med denna systematik ledde fram till en litteraturbas om 73 titlar från 2006 till 2018.

Författare	År	Titel
Brat, Denney, Giannakopoulou, Frank, Jonsson	2006	Verification of autonomous systems for space applications
Pulina, Tacchella	2010	An abstraction-refinement approach to verification of artificial neural networks
Fisher, Dennis, Webster	2013	Verifying Autonomous Systems
Szegedy, Zaremba, Sutskever, Bruna, Erhan, Goodfellow, Fergus	2013	Intriguing Properties of Neural Networks
Galko, Rossi, Savatier	2014	Vehicle-Hardware-In-The-Loop System for ADAS Prototyping and Validation
Althoff, Dolan	2014	Online Verification of Automated Road Vehicles Using Reachability Analysis
Jeannin et al	2015	A Formally Verified Hybrid System for the Next-Generation Airborne Collision Avoidance System
Patchett, Jump, Fisher	2015	Safety and Certification of Unmanned Air Systems
Seshia, Sadigh, Sastry	2015	Formal Methods for Semi-Autonomous Driving
Moosavi-Dezfooli, Fawzi, Frossard	2015	DeepFool: a simple and accurate method to fool deep neural networks
Russell, Dewey, Tegmark	2015	Research Priorities for Robust and Beneficial Artificial Intelligence
Koopman, Wagner	2016	Challenges in Autonomous Vehicle Testing and Validation
Zhao, Lam, Peng, Bao, LeBlanc, Nobukawa, Pan	2016	Accelerated Evaluation of Automated Vehicles Safety in Lane-Change Scenarios Based on Importance Sampling Techniques
Huang, Kwiatkowska, Wang, Wu	2016	Safety Verification of Deep Neural Networks
Helle, Schamai, Strobel	2016	Testing of Autonomous Systems – Challenges and Current State-of-the-Art
Papernot, McDaniel, Sinha, Wellman	2016	Towards the Science of Security and Privacy in Machine Learning
Zhao	2016	Accelerated Evaluation of Automated Vehicles (PhD thesis)
Tuncali, Pavlic, Fainekos	2016	Utilizing S-TaLiRo as an automatic test generation framework for autonomous vehicles
Harald	2016	State of the Art Software Development in the Automotive Industry and Analysis upon Applicability of Software Fault Prediction (PhD thesis)
Bastani, Ioannou, Lampropoulos et al	2016	Measuring Neural Net Robustness with Constraints
Salay, Queiroz, Czarnecki	2017	An Analysis of ISO 26262: Using Machine Learning Safely in Automotive Software
Kaur, Sobti	2017	Current challenges in modelling advanced driver assistance systems: Future trends and advancements
Burton, Gauerhof, Heinzemann (Bosch)	2017	Making the Case for Safety of Machine Learning in Highly Automated Driving
Dreossi, Donze, Seshia	2017	Compositional Falsification of Cyber-Physical Systems with Machine Learning Components
Katz, Barrett, Dill, Julian, Kochenderfer	2017	Towards Proving the Adversarial Robustness of Deep Neural Network
Abbas, O'Kelly, Rodionova, Mangharam	2017	A Driver's License Test for Driverless Vehicles
Desai, Dreossi, Seshia	2017	Combining Model Checking and Runtime Verification for Safe Robotics
Nushi, Kamar, Horvitz, Kossmann	2017	On human intellect and machine failures: Troubleshooting integrative machine learning systems
Katz, Barrett, Dill, Julian, Kochenderfer	2017	Reluplex: An efficient SMT solver for verifying deep neural networks
Dreossi, Ghosh, Sangiovanni-Vincentelli, Seshia	2017	Systematic Testing of Convolutional Neural Networks for Autonomous Driving
Seshia, Sadigh, Sastry	2017	Towards Verified Artificial Intelligence
Dreossi, Donze, Seshia	2017	Compositional Falsification of Cyber-Physical Systems with Machine Learning Components
Goodfellow, Papernot	2017	The challenge of verification and testing of machine learning
Jha, Raman, Sadigh, Seshia	2017	Safe Autonomy Under Perception Uncertainty Using Chance-Constrained Temporal Logic (earlier version called Automated Synthesis of Safe Autonomous Vehicle Control Under Perception Uncertainty)
Richter, Roy	2017	Safe Visual Navigation via Deep Learning and Novelty Detection
Rocklage, Kraft	2017	Automated Scenario Generation for Regression Testing of Autonomous Vehicles
Schuldt, Reschka, Maurer	2017	A Method for an Efficient, Systematic Test Case Generation for Advanced Driver Assistance Systems in Virtual Environments
Pei, Cao, Yang, Jana	2017	Towards Practical Verification of Machine Learning: The Case of Computer Vision Systems
Cheng, Diehl, Hamza et al.	2017	Neural Networks for Safety-Critical Applications - Challenges, Experiments and Perspectives
Guiochet, Machin, Waeselynck	2017	Safety-critical advanced robots: A survey

Pei, Cao, Yang, Jana	2017	DeepXplore: Automated Whitebox Testing of Deep Learning Systems
McAllister et al	2017	Concrete Problems for Autonomous Vehicle Safety: Advantages of Bayesian Deep Learning
Amodei, Olah, Steinhardt, Christiano, Schulman, Mane	2017	Concrete Problems in AI Safety
Rocklage	2017	Teaching self-driving cars to dream: A deeply integrated, innovative approach for solving the autonomous vehicle validation problem
Koopman, Wagner	2017	Autonomous Vehicle Safety: An Interdisciplinary Challenge
Bunel, Turkaslan, Torr, Kohli, Kumar	2017	A Unified View of Piecewise Linear Neural Network Verification
Selsam, Liang, Dill	2017	Developing Bug-Free Machine Learning Systems With Formal Mathematics
Tian, Pei, Jana, Ray	2018	DeepTest- Automated Testing of Deep-Neural-Network-driven Autonomous Cars
Gheraibia, Kabir, Djafri, Krimou	2018	An Overview of the Approaches for Automotive Safety Integrity Levels Allocation
Nushi, Kamar, Horvitz	2018	Towards Accountable AI: Hybrid Human-Machine Analyses for Characterizing System Failure
Zhang, Zhang, Zhang, Liu, Khurshid	2018	DeepRoad: GAN-based Metamorphic Autonomous Driving System Testing
Wicker, Huang, Kwiatkowska	2018	Feature-Guided Black-Box safety testing of Deep Neural Networks
Johnson	2018	The Increasing Risks of Risk Assessment: On the Rise of Artificial Intelligence and NonDeterminism in Safety-Critical Systems
Winikoff	2018	Towards Trusting Autonomous Systems
Rabatin, Voros	2018	Towards the Verification of Neural Networks for Critical Cyber-Physical Systems
Sun, Huang, Kroening	2018	Testing Deep Neural Networks
Krakovna, Orseau, Martic, Legg	2018	Measuring and avoiding side effects using relative reachability
Dreossi, Jha, Seshia	2018	Semantic Adversarial Deep Learning
Schwarting, Alonso-Mora, Rus	2018	Planning and Decision-Making for Autonomous Vehicles (Section 5. Verification and synthesis)
Kuper, Katz, Gottschilich, Julian, Barrett, Kochenderfer	2018	Toward Scalable Verification for Safety-Critical Deep Networks
Brundage, Avin et al.	2018	The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation
Victoria Krakovna	2018	Deep Safety Blog - AI safety resources:
Mason, Calinescu, Kudenko, Banks	2018	Assurance in Reinforcement Learning Using Quantitative Verification
Timperley, Afzal, Katz, Hernandez, Le Goues	2018	Crashing simulated planes is cheap: Can simulation detect robotics bugs early?
Koopman, Wagner	2018	Toward a Framework for Highly Automated Vehicle Safety Validation
Ruan, Huang, Kwiatkowska	2018	Reachability analysis of Deep Neural Networks with Provable guarantees
Claessen, Smallbone, Eddeland, Ramezani, Akesson	2018	Using Valued Booleans to Find Simpler Counterexamples in Random Testing of Cyber-Physical Systems
Sun, Wu, Ruan, Huang, Kwiatkowska, Kroening	2018	Concolic Testing for Deep Neural Networks
Goodfellow, McDaniel, Papernot	2018	Making Machine Learning Robust Against Adversarial Inputs
Ma, Zhang, Xue, Li, Liu, Zhao, Wang	2018	Combinatorial Testing for Deep Learning Systems
Gopinath, Katz, Păsăreanu, Barrett	2018	DeepSafe: A Data-Driven Approach for Assessing Robustness of Neural Networks

Tabell 2. Sammanställning av teoretiska källor analyserade i **FramTEST**

För att analysera litteraturmaterialet tillämpade **FramTEST** en systematisk metod för att temabaserat genomföra en litteraturgranskning (Wolfwinkel et al. 2013). Metoden syftar till att genom forskningsmetoden grundad teori vägleda forskare att från en litteraturstudie vaska fram möjliga faktorer exempelvis under en förstudie som därefter kan verifieras i en realiseringsstudie. Modellen menar att teoretiska källor kan betraktas som grund för att identifiera och koda faktorer vilka därefter grupperas till kategorier och övergripande teman.

Materialet som analyserades generade en central insikt för förstudiens fortsättning. Över åren har omfattande kunskap utformats kring krav och behov på verktyg, metod och standarder. Emellertid finns en kunskapslucka vad gäller insikter hur ny testmetodik som omformar testning bör implementeras. Med andra ord förstudiens inledande inriktning mot hur testmetodik utformas, vidgades till att fånga hur nya förutsättningar för testning implementeras.

Arbetet medförde att 12 faktorer kunde identifieras vilka grupperades till tre övergripande kategorier (mekanismer) som i samspel har betydelse när en verksamhet ska ställa om hur testning bedrivs. De tre övergripande kategorier som med tillhörande faktorer definierades blev:

- Kategori 1: organisationens *tillit och övertygelse* för hur testning utförs
- Kategori 2: organisationens *processer och rutiner* för hur testning utförs
- Kategori 3: organisationens *verktyg och infrastruktur* för hur testning utförs

I det teoretiska materialet, såväl som i förgående projekt inom EMK är processer, rutiner, verktyg och infrastruktur för testning framträdande. I senare artiklar pekar materialet emellertid på att förändring mot datadriven, AI och ML-baserad, testning förutsätter att organisationens övertygelse och tillit på hur testning ska utföras justeras. Samspelet mellan socio-tekniska faktorer är i sig inget

nytt. Garud & Rappa (1994) talar till exempel om vikten av att förstå övertygelse, rutiner och artefakt när innovation och organisering studeras. Vad som är nytt är applikationsområdet testning samt de underliggande förslag på faktorer som analysen i **FramTEST** genererat.

Resultatet av litteraturstudien skapade förutsättningar för att inom **FramTEST** 1) förbereda intervjustudien, 2) format basen för ett artikelmanus vilket efter förstudien under 2019 kommer färdigställas och skickas in för granskning till forskningskonferensen HICSS53 (se avsnitt 6.2).

5.2 Fas 2: explorativ intervjustudie

Resultatet av litteraturstudien skapade förutsättningar för att utforma grundstrukturen för en forskningsmodell att förstå och kunna förklara vilka mekanismer (faktorer) som påverkar implementeringen av ny testmetodik inom fordonsindustrin, där centrala komponenter skiljer mot den logik som testning nuvarande processer tillämpar. De nya komponenter som är i fokus för **FramTEST** är ML, AI och automatisering. Analysen av litteraturen genererade 12 förslag på faktorer som grupperades i tre övergripande socio-tekniska kategorier (se avsnitt 5.1). Litteraturgenomgången skapade främst grund för kategori 2 och 3: process och rutin, samt verktyg och infrastruktur. Kategori 1 som fokuserar organisationens förtroende och uppfattning kring testarbetets logik berördes i mindre omfattning i den litteratur som granskningen omfattat. En möjlig förklaring till detta är att forskningen inom området primärt fokuserat den tekniska dimensionen av hur testning ska bedrivas inom fordonsindustrin. Intressant att se dock är att socio-dimensionen lyfts fram som potentiellt hinder i teoretiska källor av nyare årgång. Således ny metodik, nya verktyg och ny logik för testning kan tas fram, men om inte organisationen anammar en omvälvning genom de tekniska möjligheter som det nya stöder medför, kommer ej utväxling ske av hur testning utförs. För att stärka den teoretiskt grundade forskningsmodellen samt validera relevansen i de tre kategorier som modellen omfattar (med tillhörande faktorer) utformades under sensommaren 2018 en intervjuguide som grund för att genomföra en explorativ intervjustudie med experter från testning inom fordonsindustrin samt närliggande branscher. Med närliggande branscher menas i **FramTEST** organisationer som antingen levererar tjänster till fordonsindustrin eller produkter och delkomponenter.

ID	Roll	Erfarenhet/år	Typ av företag	Karaktärsdrag på verksamhet	Bransch
Respondent Ö	Testledare	19	Incumbent	Utländskt företag som investerar i svenskt kunnande inom fordonsbranschen	Fordonsindustrin
Respondent Ä	Testledare	15	Scaleup	Ny startat joint venture mellan aktörer med lång erfarenhet inom fordonsindustrin	Fordonsindustrin
Respondent Å	Expert test och verifiering	9	Incumbent	Global fordonstillverkare	Fordonsindustrin
Respondent X	Testare	13	Incumbent	Före detta statligt företag med rötter inom Telecom och mediadistribution.	Leverantör till fordonsindustrin
Respondent Y	Testare	20	Scaleup	Ungt bolag som har en tydlig nisch med sina främsta kunder utomlands (USA).	Leverantör till fordonsindustrin
Respondent Z	Testledare	20	Incumbent	Ett svenskt storbolag som främst verkar inom banbunden logistik och fordonstillverkning.	Fordonsindustrin
Respondent W	Testledare	5	Incumbent	Dotterbolag inom stor fordonskoncern.	Fordonsindustrin
Respondent V	Testledare	17	Incumbent	Västsvenskt företag med fokus på att hjälpa små och medelstora företag med validering, verifiering och revision.	Leverantör till fordonsindustrin
Repåndent M	Testare	7	Incumbent	Dotterbolag till stor fordonstillverkare i Västsverige.	Leverantör till fordonsindustrin
Respondent N	Expert test och verifiering	18	Incumbent	Global fordonstillverkare	Fordonsindustrin
Respondent T	Expert test och verifiering	20	Incumbent	Forskningsleverantör till fordonsindustrin och samhället	Angränsande bransch
Respondent S	Expert test och verifiering	12	Scaleup	Teknikkonsultbolag med fokus på testning och kvalitetsarbete.	Leverantör till fordonsindustrin

Tabell 3. Sammanställning av respondenter intervjuade i **FramTEST**

Sexton möjliga respondenter identifierades från lika många organisationer. I urvalet utgick forskarteamet från de relationer och kontakter som teamet har. Kontakt togs även med forskarlaget

i SMILE II för att inledningsvis synka relationer inom industrin, i andra hand, tillsammans utforma intervjuguiden. Indelningen av de potentiella respondenterna gjordes utifrån tre kriterier: roll, antal år inom testning och typ av företag. För att säkerställa koppling till operativ testning inkluderades respondenter som i vardagen arbetar med testning inom olika områden (testare). För att säkerställa koppling till en taktisk nivå där organisering av testarbete utförs valdes respondenter ut som leder testarbete eller ansvarar för olika testteam med koppling till fordonsindustrin. För att säkerställa långsiktig utblick och vision kopplat till testning valdes även ett antal respondenter som är experter inom området med nödvändigtvis inte är en del av det operativa arbetet med testning eller ledning av detsamma. Av sexton respondenter valde tolv att under hösten ställa upp på intervjuer. Profilerna på deltagandet respondenter beskrivs i Tabell 3.

Med utgångspunkt i resultatet från state-of-art-fasen arrangerades tre arbetsmöten med representanter från det pågående SMILE II. Tillsammans utformades den intervjuguide som användes för att genomföra de tolv intervjuerna. Motivet till att samverka med SMILE II forskare var tvåfaldigt; dels den nära kopplingen i ambition mellan FramTEST som förstudie och SMILE II som forskningsprojekt, dels möjligheten att stärka SMILE II genom att korsbefrukta de resultat som generades i FramTEST med de resultat som utvecklats inom SMILE och kommer utvecklas inom SMILE II.

Tema	Frågor
Inledning	Hur mycket erfarenhet har du av att arbeta med ML, AI och automatisering inom test?
	Hur mycket erfarenhet har du av att jobba med säkerhetsstandarder vid test?
	Hur mycket erfarenhet har du av att jobba med testning?
	Beskriv din nuvarande roll?
Nuläge	Var i produktutvecklingscykeln kommer ni in?
	Vilka metoder och standarder använder ni idag?
	Hur oberoende är ni jämfört med utvecklarna? (separerade eller blandade)
	Hur stort är ert testteam relativt utvecklarna? (ratio)
	Upplever du att ni hinner med ert ansvarsområde idag?
Önskat läge	Upplever du att ni är prioriterade i er organisation idag?
	Var i produktutvecklingscykeln vill ni komma in?
	Vilka arbetssätt vill ni använda?
	Hur stora team vill ni vara?
	Hur oberoende vill ni vara jämfört med utvecklarna?
	Vilka verktyg och gränssnitt vill ni använda?
	Vilka kompetenser vill ni ha tillgång till?
Hinder	Vilka metoder och standarder vill ni jobba med?
	Vilka värden kan ML, AI och Automatisering medföra på testverksamheten?
	Vilka utmaningar ser du för komma till det önskade läget?
Förutsättningar	Vilka är orsakerna till hindren?
	Finns kunskaper för att förändra hur ni utför test/förståelse för behovet?
	Saknar ni rutiner, ramverk eller standarder?
Aktiviteter	Har ni lämpliga verktyg och data?
	Vilka är nästa steg att nå det önskade läget
Avslutning	Hur ser förändringstakten ut för att ställa om hur ni verifierar och validerar?
	Beskriv hur du ser på logiken att förändra testning - vad ska man ändra först – förståelse, rutin eller verktyg?
	Är det något som vi inte frågat om som du ser som relevant för testning i framtiden inom fordonsindustrin?

Tabell 3. Intervjuguide tillämpad i FramTEST

Intervjuguiden utformades genom sju teman. En *inledning* för att kunna positionera respondenten i förhållande till övriga respondenter och förstudiens syfte. Guiden omfattade även två avslutande teman, *aktiviteter* och *avslutning*, för att fånga och diskutera hur den aktuella organisationen arbetar med att utveckla på det sätt som testning bedrivs i relation till de förändrade förutsättningar som ML, AI och Automatisering medför, samt det önskade läge som intervjun dessförinnan behandlat. Däremellan genom fyra teman, *nuläge*, *önskat läge*, *hinder* och *förutsättningar*, ämnade intervjuguiden stimulera till en semi-strukturerad dialog mellan intervjuerna och respondenten om vad som påverkar en organisation att ändra hur testning bedrivs. De faktorer som identifierats utgjorde här grund för följdfrågor och diskussionspunkter under den dialog som intervjuerna skapade. Genom öppenheten i frågorna skapades förutsättningar att respondenterna berörde faktorer som ej identifierats genom litteraturgranskningen, samt berika förståelsen och innebörden i de som utgjorde utgångsläge för intervjuerna. I några av fallen så tillhandahöll respondenter även dokumentation som t.ex. beskrev nuvarande process för testning, rollbeskrivningar samt illustration av testmetodik.

Datakälla	Datainsamlingar	Volym
	<ul style="list-style-type: none"> Intervjuer med aktörer inom fordonsindustrin (6) <ul style="list-style-type: none"> - Testledare (4) - Expert (visionär) (2) 	12 Semi-strukturerade Intervjuer 12 Respondenter Längd (minuter): medel = 58 min Antal transkriberade ord: 103 950 ord
Intervjuer	<ul style="list-style-type: none"> Intervjuer med aktörer som utgör leverantörer eller utgör angränsande områden till fordonsindustrin (6) <ul style="list-style-type: none"> - Testare (3) - Testledare (1) - Expert (visionär) (2) 	
Dokument	Processbeskrivningar, metodbeskrivningar samt rollbeskrivningar	Antal källor: 9

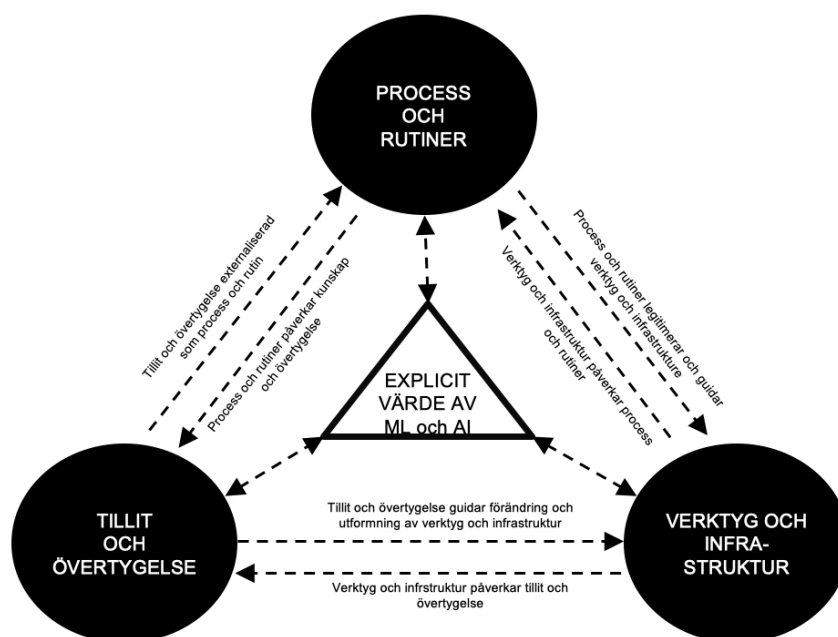
Intervjuerna genomfördes av två personer från forskningsteamet. Ljudupptagningarna transkriberades i efterhand vilket genererade ett rikt kvalitativt datamaterial som omfattar 110 sidor intervjuutskrift omfattande knappt 104 000 ord.

5.3 Fas 3: analys och syntes

Den tredje fasen i förstudien som i kalendertid omfattade november och december 2018 riktades in mot att analysera det insamlade och transkriberade intervjumaterialet och relatera detta till state-of-art. Arbetet utfördes genom två workshops med projektteamet samt att en av forskarna påbörjade arbetet att koda datamaterialet genom analysverktyget atlas.ti. Kontakt togs även via en upparbetad relation med dataanalysspecialisten SAS Institute med frågan att få ta del av deras automatiserade textanalysverktyg för att automatiskt gruppera materialet i förhållande till olika kategorier. Inom ramen för förstudien kunde inte detta genomföras men under våren när materialet ytterligare kommer att analyseras är förhoppningen att atlas.ti kan kompletteras med automatiserad textanalys.

Resultatet av den tematiska kvalitativ analys som genomfördes i fas 3 blev att ytterligare fem förslag på faktorer identifierades. Materialet stärker dessutom de tolv faktorer som identifierades från state-of-art stärktes. Forskarteamet bedömer att i materialet även besitter grund för att sätta upp hypoteser där några faktorer utgör beroende och andra oberoende variabler. På så sätt kan fortsatt forskning studera hur olika mekanismer påverkar förutsättningarna att implementera nya förutsättningar för hur testning ska ske inom olika faser vid utveckling mjukvara för fordonssystem.

Utöver detta identifierades vad som har potential att förklaras som en övergripande beroende variabel. Återkommande i intervjuerna beskrivs att en ny logik för testning förutsätter ett samspel mellan tillit och övertygelse, process och rutin, samt verktyg och infrastruktur för testning. Detta samspel är beroende av att det värdet med en ny logik för testning som t.ex. baseras på ML, AI och högre grad av automatisering är explicit och uppenbar för organisationen där den nya testmetodiken ska implementeras. Ett möjligt samspel mellan de socio-tekniska faktorerna ligger i linje med Lanzara & Patriotta (2007) som i en studie av förändringar av arbetssätt inom fordonsindustrin beskriver värdet av en "template" som vägleder organisationen att förstå värdet med förändringen. En template definieras som "*a practical example, often based on a shared cognitive analogy, which, because of its ontological 'obviousness', soon becomes the accepted way of doing things — a master model or pattern*" (Lanzara & Patriotta, 2007) En spännande utmaning är formulera och skapa template underlättar införandet av en ny testlogik där densamma är materialiserad i verktyg, infrastruktur, process och rutiner, och därmed stärker tillit och övertygelse för det nya sättet att testa mjukvara – baserat på komponenter som innefattar ML, AI och automatisering.



Figur 3. Övergripande modell för hur förstå och förklara hur implementering av testmetodik kan genomföras som möter framtidens krav på verifierade och validerade fordonsystem

Figur 3 beskriver på övergripande nivå den forskningsmodell som utformats och grundats i teori och empiri genom förstudien **FramTEST**. Modellen och dess underliggande faktorer ligger till grund för ett andra manuskript på vetenskaplig rapport som under våren 2019 kommer färdigställas (se avsnitt 6.2). Artikeln kommer presentera ett antal hypoteser som kopplas till faktorerna och det är forsknings-teamets ambition att inom ramen för pågående forskning eller framtida projekt inom EMK kunna tillämpa modellen för att förstå och förklara hur testning inom fordonsindustrin kan förändras mot högre grad av automatisering samt tillämpning av ML och AI. Av särskilt intresse är att förstå vad som kan utgöra en generative template som förmedlar explicit och uppenbart värde till organisationen så att tillit och övertygelse byggs kring datadrivna processer och rutiner genom förändrade verktyg och anpassad infrastruktur för testning av mjukvara inom fordonsindustrin.

6 Spridning och publicering

Under förstudien har kunskapsspridning genomförts samt förberetts i två avseenden. Kunskapsspridningsseminarier som en del av de återkommande frukostseminarium som Infotiv AB arrangerar, dels genom förberedandet av två vetenskapliga manuskript som vilar på det arbete som förstudien utfört. En tredje insats är det samarbete som etablerades mellan **FramTEST** och det pågående forskningsprojektet SMILE II.

6.1 Kunskaps- och resultatspridning

Hur har/planeras projektresultatet att användas och spridas?	Markera med X	Kommentar
Öka kunskapen inom området	X	Förstudien har generat förutsättningar för vetenskaplig publicering samt ytterligare kunskapsutveckling genom att utgöra input till pågående projekt inom FFI/EMK samt ingångsvärde som del av nytt projekt.
Föras vidare till andra avancerade tekniska utvecklingsprojekt	X	Se ovan

Föras vidare till produktutvecklingsprojekt	X	Se ovan
Introduceras på marknaden		Resultatet är ej av denna karaktär
Användas i utredningar/regelverk/tillståndsärenden/politiska beslut	X	Resultatet av förstudien och förstående artiklar har ett värde för framtida utredningar och forskning hur ny logik för testning implementeras i fordonsindustrin.

Progressionen i projektet har presenterats vid två tillfällen under kunskapsspridningsevent som Infotiv arrangerar: Den 16 maj 2018 då ett seminarium arrangerades kring Högremsjukvarukvalitet, med Richard Torkar och Robert Feldt som keynotes. Den 22 november vid ett seminarium på temat Model-based test methods for cyber-physical systems. Medverkande organisationer på seminarierna har varit Volvo AB, Volvo Cars, Volvo GTT, Volvo Buss, Volvo Penta, CEVT, T-Engineering, Bombardier, Ascom, Nibe, Q-Matic, ÅF, Etraveli, SAAB Surveillance, ESAB, Mitsubishi electric, Zenuity, C-Pac, Recorded Future, Collector Bank, VCRS, Zenuity, Seal Software, Kollmorgen, Sigma IT, House of test och Quality Mind.

6.2 Publikationer

Två artikelmanus har planerats från projektet och ambitionen är att under 2019 färdigställa och skicka in dessa vetenskapliga bidrag för granskning till vetenskaplig konferens och journal. En risk här är att färdigställandet saknar finansiering:

Artikelmanus 1: *“Towards Data-driven Verification and Validation of Autonomous Transport Systems: An Literature Review”* för konferensen *Hawaiian Conference of Systems Science 53* Deadline 15 juni 2019

Artikelmanus 2: *“Building a Data-driven Validation Culture: A Framework for how Automotive Firms can Transform the Practice of Software Testing”* till tidskriften *The Journal of Strategic Information Systems*

Utöver dessa publikationer så korsbefruktas nu data insamlat genom SMILE II med resultatet från **FramTEST**. Målet är att vidareförandet av datan in i SMILE II kan medföra i ytterligare en till två artiklar samförfattade av forskare från båda forskningsteam.

7 Slutsatser och fortsatt forskning

Förstudien **FramTEST** har efter state-of-art genomgången arbetat med forskningsfrågan *Hur bör testmetodik implementeras som möter framtidens krav på verifierade och validerade fordonssystem?*

Slutsatsen är verifiering och validering av framtida fordonssystem kräver att testning genomförs på delvis andra sätt som kräver nya komponenter baserade på ML, AI och automatisering. En utmaning är att skapa ny testmetodik, men förstudiens genomlysning av forskning från FFI/EMK samt litteratur inom testning för fordonsindustrin visar på omfattande forskning om sådan design. Istället har förstudien fångat och konceptualiserat den framtida utmaningen att implementera en ny logik för testning. För att lyckas med det blir antagandet från **FramTEST** att en ny logik kräver en socio-teknisk omvälvning av det sätt som testning bedrivs. Förstudien föreslår en forskningsmodell för att studera de mekanismer som behöver samspela för att bygga en datadriven testprocess som baseras intelligenta komponenter.

8 Deltagande parter och kontaktpersoner

Förstudien **FramTEST** har genomförts av ett forskningsteam som omfattat forskare från RISE Viktoria och samt testexperter från affärsområdet test och validering vid Infotiv AB.



Anders Hjalmarsson Jordanius, Projektledare
Ana Magazinius, Forskare
Alexey Voronov, Forskare
Cristofer Englund, Forskningsledare
Patrik Andersson, Forskare

Hanna Gynnerstedt, Konsultchef
Martin Karsberg, Testexpert
Cecilia Kjellman, Testexpert
Tomas Sjögren, Affärsområdesansvarig Test
och validering

Primär kontaktperson för **FramTEST** är: Anders Hjalmarsson Jordanius
email: anders.hjalmarsson@ri.se
Telefon: 0707567870